GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 05:58:19 ; Search time 2471.47 Seconds
(without alignments)
4083.506 Million cell updates/sec

Title: US-09-294-663-2
Perfect score: 2821
Sequence: 1 GAAAAGAATAACCAGCGAAC.....TAAAAAAAAAAAAAAAAAAA 2821

Scoring table: IDENTITY_NUC

Searched: 808301 seqs, 1788773984 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl1:*
9: gb_pl2:*
10: gb_pr1:*
11: gb_pr2:*
12: gb_pr3:*
13: gb_ro:*
14: gb_st:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: gb_htg:*
19: em_ba:*
20: em_fun:*
21: em_hum1:*
22: em_hum2:*
23: em_in:*
24: em_om:*
25: em_or:*
26: em_ov:*
27: em_pat:*
28: em_ph:*
29: em_pl:*
30: em_ro:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: em_htg:*
35: em_sts:*
36: gb_ba1:*
37: gb_ba2:*
38: gb_pl1:*
39: gb_pl2:*
40: gb_pr1:*
41: gb_pr2:*
42: gb_pr3:*
43: gb_sts:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2821	100.0	2821	3	TNAF000606	AF000606 Trichoplu
2	2354.4	83.5	2455	3	TNAF000605	AF000605 Trichoplu
3	198	7.0	15720	11	HUMMUC2X	L21998 Homo sapien
4	198	7.0	15720	41	HUMMUC2X	L21998 Homo sapien
c 5	166.4	5.9	185994	18	AC002042	AC002042 *** SEQUE
c 6	158.8	5.6	10596	6	I25041	I25041 Sequence 15
c 7	158.8	5.6	10596	6	I30503	I30503 Sequence 15
c 8	158.8	5.6	5452	15	U02454	U02454 Cloning vec
c 9	158.8	5.6	10850	15	U02455	U02455 Cloning vec
c 10	158.8	5.6	10737	15	XXU02428	U02428 Cloning vec
c 11	158.8	5.6	172281	17	EBV	V01555 Epstein-Bar
c 12	158.8	5.6	184113	17	HS4B958RAJ	M80517 Epstein-Bar
c 13	158.8	5.6	1150	17	HS4ULIR3	J02079 epstein-bar
14	149.4	5.3	3173	17	D88734	D88734 Equine herp
15	146.8	5.2	2070	5	BSU43200	U43200 Boreogadus
16	146.2	5.2	3170	17	D88733	D88733 Equine herp
17	140.2	5.0	1720	10	HUMMUC2A	M74027 Human mucin
18	140.2	5.0	1720	40	HUMMUC2A	M74027 Human mucin
c 19	138	4.9	3941	13	RNLAR2	X83546 R.norvegicu
20	137	4.9	3811	10	HUMIMUCA	M94131 Human mucin
21	137	4.9	3811	40	HUMIMUCA	M94131 Human mucin
22	136.2	4.8	33359	3	CEF55B11	Z83318 Caenorhabdi
c 23	134.6	4.8	2479	13	RNU87960	U87960 Rattus norv
c 24	128.4	4.6	28291	3	CEF56H9	Z74473 Caenorhabdi
25	124.8	4.4	1046	5	PSUAFA02	M63477 P.americanu
c 26	122.8	4.4	174428	12	AC005821	AC005821 Homo sapi
c 27	122.8	4.4	174428	42	AC005821	AC005821 Homo sapi
c 28	122.4	4.3	123110	12	AC005383	AC005383 Homo sapi
c 29	122.4	4.3	123110	42	AC005383	AC005383 Homo sapi
30	121.6	4.3	5719	13	MMCATS1	AF051726 Mus muscu
c 31	121.4	4.3	48000	18	AC005930	AC005930 *** SEQUE
c 32	120.2	4.3	2793	6	AR009990	AR009990 Sequence
c 33	120.2	4.3	2793	6	I95876	I95876 Sequence 1
c 34	119.2	4.2	3489	17	KSU52064	U52064 Kaposi's sa
35	118.8	4.2	3834	5	DMU43149	U43149 Dissostichu
c 36	118.8	4.2	4635	12	HSA6805	AJ006805 Homo sapi
c 37	118.8	4.2	4635	42	HSA6805	AJ006805 Homo sapi
c 38	118.6	4.2	3060	3	AF027735	AF027735 Nephila c
c 39	118.6	4.2	11343	13	MMDSPPG	AJ002141 Mus muscu
c 40	118.6	4.2	4420	13	U67916	U67916 Mus musculu
41	118.4	4.2	2065	8	BLYHORDCA	M36941 Hordeum vul
42	118.4	4.2	2065	38	BLYHORDCA	M36941 Hordeum vul
43	118	4.2	5004	5	DMU58944	U58944 Dissostichu
c 44	117	4.1	1934	8	PHGCR1	X04335 Petunia grp
c 45	117	4.1	988	10	HUMIR3EBA	M17919 Human DNA w

ALIGNMENTS

RESULT 1
TNAF000606
LOCUS TNAF000606 2821 bp mRNA INV 28-JUN-1997
DEFINITION Trichoplusia ni insect intestinal mucin IIM22 mRNA, complete cds.
ACCESSION AF000606
NID g2224920
KEYWORDS .
SOURCE cabbage looper.
ORGANISM Trichoplusia ni
Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Noctuoidea; Noctuidae; Trichoplusia.
REFERENCE 1 (bases 1 to 2821)
AUTHORS Wang,P. and Granados,R.R.
TITLE Molecular cloning and sequencing of a novel invertebrate intestinal mucin cDNA
JOURNAL J. Biol. Chem. 272 (26), 16663-16669 (1997)
MEDLINE 97341213
REFERENCE 2 (bases 1 to 2821)
AUTHORS Wang,P. and Granados,R.R.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1997) Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA

```
FEATURES             Location/Qualifiers
     source          1. .2821
                     /organism="Trichoplusia ni"
                     /db_xref="taxon:7111"
     CDS             26. .2449
                     /codon_start=1
                     /product="insect intestinal mucin IIM22"
                     /db_xref="PID:g2224921"
                     /translation="MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWL
                     LPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQA
                     PATTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQA
                     PTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTP
                     AATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGY
                     TFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPA
                     NFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGES
                     EEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA
                     RPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPT
                     AAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPT
                     AAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP
                     APNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPT
                     VAPAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHADCDKYWV
                     CDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMSESYEGVQVFIPWNKLDED
                     IRQALNFEL"
BASE COUNT      710 a   1030 c    529 g    552 t
ORIGIN
```

```
  Query Match             100.0%;  Score 2821;  DB 3;  Length 2821;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2821;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60

Qy      61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120

Qy     121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180

Qy     181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240

Qy     241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300

Qy     301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360

Qy     361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420

Qy     421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480

Qy     481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540

Qy     541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600

Qy     601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660

Qy     661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720

Qy     721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780

Qy     781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840

Qy     841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900

Qy     901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960

Qy     961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020

Qy    1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080

Qy    1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140

Qy    1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200

Qy    1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260

Qy    1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320

Qy    1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380

Qy    1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440

Qy    1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500

Qy    1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560

Qy    1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620

Qy    1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680

Qy    1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740
```

```
Qy   1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800

Qy   1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860

Qy   1861 ATCACCACCTACTGCTGCCCCCCACTACCGCAGCACCTGCCCCCCAACACCACAGTCACTGT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1920

Qy   1921 ACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980

Qy   1981 ACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCCAACACCACAGTCACTGTACCACC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 ACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCCAACACCACAGTCACTGTACCACC 2040

Qy   2041 CACTGCTGCCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100

Qy   2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160

Qy   2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220

Qy   2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280

Qy   2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340

Qy   2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400

Qy   2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460

Qy   2461 TTAATGAAGGTTTTGTTTTATTTTTGAGTTATTATTCCAATGGGCGGGAAAGTCCGCCAT 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TTAATGAAGGTTTTGTTTTATTTTTGAGTTATTATTCCAATGGGCGGGAAAGTCCGCCAT 2520

Qy   2521 TATTGGGTCTTGCCAGTTTTGAGGAAACCTTTTTTTTTTACTACCAACATTCTTGTGAACC 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 TATTGGGTCTTGCCAGTTTTGAGGAAACCTTTTTTTTTTACTACCAACATTCTTGTGAACC 2580

Qy   2581 CATATTTATTACGTATTAAACATCGTGATTTGAAAAACGTTACATGATTTTTTCATTAAT 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CATATTTATTACGTATTAAACATCGTGATTTGAAAAACGTTACATGATTTTTTCATTAAT 2640

Qy   2641 TTGAAACAATTTATGTTGTTTTTGTTCTCATTAAATATCAAATATCATTTTCGAAACTGG 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 TTGAAACAATTTATGTTGTTTTTGTTCTCATTAAATATCAAATATCATTTTCGAAACTGG 2700

Qy   2701 CAATTTTGGATTGGAATAATCAACAAATGGTTAAGAAAAAAAAACGATTTCTTAAAAATGT 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CAATTTTGGATTGGAATAATCAACAAATGGTTAAGAAAAAAAAACGATTTCTTAAAAATGT 2760

Qy   2761 ATTTATTATAAAATGTGTAAATAAATATACAAATTAGCATTTAAAAAAAAAAAAAAAAAA 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 ATTTATTATAAAATGTGTAAATAAATATACAAATTAGCATTTAAAAAAAAAAAAAAAAAA 2820
```

```
Qy   2821 A 2821
          |
Db   2821 A 2821
```

```
RESULT   2
TNAF000605
LOCUS       TNAF000605   2455 bp    mRNA            INV       28-JUN-1997
DEFINITION  Trichoplusia ni insect intestinal mucin IIM14 mRNA, complete cds.
ACCESSION   AF000605
NID         g2224918
KEYWORDS    .
SOURCE      cabbage looper.
  ORGANISM  Trichoplusia ni
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Noctuoidea;
            Noctuidae; Trichoplusia.
REFERENCE   1  (bases 1 to 2455)
  AUTHORS   Wang,P. and Granados,R.R.
  TITLE     Molecular cloning and sequencing of a novel invertebrate intestinal
            mucin cDNA
  JOURNAL   J. Biol. Chem. 272 (26), 16663-16669 (1997)
  MEDLINE   97341213
REFERENCE   2  (bases 1 to 2455)
  AUTHORS   Wang,P. and Granados,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1997) Boyce Thompson Institute, Cornell
            University, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .2455
                     /organism="Trichoplusia ni"
                     /db_xref="taxon:7111"
     CDS             38. .2404
                     /codon_start=1
                     /product="insect intestinal mucin IIM14"
                     /db_xref="PID:g2224919"
                     /translation="MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWL
                     LPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQA
                     PATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQA
                     PTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTP
                     AATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGY
                     TFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPA
                     NFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGES
                     EEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA
                     RPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPT
                     AAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPT
                     AAPTTAVPEIPITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP
                     APNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPA
                     TTPEDDDIDPPLPNDPINPCVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNP
                     TTKTCDFACNVGCVRSNIQMSESYEGVQVFIPWNKLDEDIRQALNFEL"
BASE COUNT      592 a    961 c    479 g    423 t
ORIGIN
```

```
  Query Match             83.5%;  Score 2354.4;  DB 3;  Length 2455;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 2422;  Conservative    0;  Mismatches    1;  Indels   57;  Gaps    1;

Qy       1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 72

Qy      61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 132

Qy     121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 192

Qy     181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     193 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 252
```

```
Qy    241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 312

Qy    301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    313 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 372

Qy    361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    373 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 432

Qy    421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 492

Qy    481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    493 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 552

Qy    541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    553 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 612

Qy    601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    613 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 672

Qy    661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    673 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 732

Qy    721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    733 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 792

Qy    781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    793 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 852

Qy    841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    853 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 912

Qy    901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    913 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 972

Qy    961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    973 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1032

Qy   1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1033 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1092

Qy   1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1093 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1152

Qy   1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1153 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1212

Qy   1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1213 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1272

Qy   1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1273 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1332

Qy   1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1333 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1392

Qy   1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1393 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1452

Qy   1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1453 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1512

Qy   1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1513 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1572

Qy   1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1573 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1632

Qy   1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1633 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1692

Qy   1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1693 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1752

Qy   1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1753 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1812

Qy   1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1813 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1872

Qy   1861 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1873 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1932

Qy   1921 ACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980
          ||||||||||||
Db   1933 ACCACCCACTGCT---------------------------------------------- 1946

Qy   1981 ACCCACCGCAGCCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 2040
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1946 ----------GCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 1995

Qy   2041 CACTGCTGCCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1996 CACTGCTGCCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCAACACCACAGCTGC 2055

Qy   2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2056 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2115

Qy   2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2116 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2175

Qy   2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2176 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2235

Qy   2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2236 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2295

Qy   2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2296 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2355

Qy   2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2356 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2415

Qy    2461 TTAATGAAGGTTTTGTTTTA 2480
          ||||||||||||||||||||
Db    2416 TTAATGAAGGTTTTGTTTTA 2435


RESULT   3
HUMMUC2X
LOCUS       HUMMUC2X    15720 bp    mRNA            PRI       09-JAN-1998
DEFINITION  Homo sapiens intestinal mucin (MUC2) mRNA, complete cds.
ACCESSION   L21998
NID         g454153
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 15720)
  AUTHORS   Gum,J.R.Jr. ., Hicks,J.W., Toribara,N.W., Kim,Y.S. and Siddiki,B.
  TITLE     Molecular cloning of human intestinal mucin (MUC2) cDNA:
            Identification of the amino terminal and overall sequence
            similarity to pre-pro-von Willebrand factor
  JOURNAL   J. Biol. Chem. 269, 2440-2446 (1994)
  MEDLINE   94132002
REFERENCE   2  (bases 1 to 15720)
  AUTHORS   Eckhardt,A.E., Timpte,C.S., DeLuca,A.W. and Hill,R.L.
  TITLE     The complete cDNA sequence and structural polymorphism of the
            polypeptide chain of porcine submaxillary mucin
  JOURNAL   J. Biol. Chem. 272 (52), 33204-33210 (1997)
  MEDLINE   98070526
FEATURES             Location/Qualifiers
     source          1. .15720
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="intestine"
     CDS             28. .15567
                     /codon_start=1
                     /product="mucin"
                     /db_xref="PID:g454154"
                     /translation="MGLPLARLAAVCLALSLAGGSELQTEGRTRYHGRNVCSTWGNFH
                     YKTFDGDVFRFPGLCDYNFASDCRGSYKEFAVHLKRGPGQAEAPAGVESILLTIKDDT
                     IYLTRHLAVLNGAVVSTPHYSPGLLIEKSDAYTKVYSRAGLTLMWNREDALMLELDTK
                     FRNHTCGLCGDYNGLQSYSEFLSDGVLFSPLEFGNMQKINQPDVVCEDPEEEVAPASC
                     SEHRAECERLLTAEAFADCQDLVPLEPYLRACQQDRCRCPGGDTCVCSTVAEFSRQCS
                     HAGGRPGNWRTATLCPKTCPGNLVYLESGSPCMDTCSHLEVSSLCEEHRMDGCFCPEG
                     TVYDDIGDSGCVPVSQCHCRLHGHLYTPGQEITNDCEQCVCNAGRWVCKDLPCPGTCA
                     LEGGSHITTFDGKTYTFHGDCYYVLAKGDHNDSYALLGELAPCGSTDKQTCLKTVVLL
                     ADKKKNAVVFKSDGSVLLNQLQVNLPHVTASFSVFRPSSYHIMVSMAIGVRLQVQLAP
                     VMQLFVTLDQASQGQVQGLCGNFNGLEGDDFKTASGLVEATGAGFANTWKAQSTCHDK
                     LDWLDDPCSLNIESANYAEHWCSLLKKTETPFGRCHSAVDPAEYYKRCKYDTCNCQNN
                     EDCLCAALSSYARACTAKGVMLWGWREHVCNKDVGSCPNSQVFLYNLTTCQQTCRSLS
                     EADSHCLEGFAPVDGCGCPDHTFLDEKGRCVPLAKCSCYHRGLYLEAGDVVVRQEERC
                     VCRDGRLHCRQIRLIGQSCTAPKIHMDCSNLTALATSKPRALSCQTLAAGYYHTECVS
                     GCVCPDGLMDDGRGGCVVEKECPCVHNNDLYSSGAKIKVDCNTCTCKRGRWVCTQAVC
                     HGTCSIYGSGHYITFDGKYYDFDGHCSYVAVQDYCGQNSSLGSFSIITENVPCGTTGV
                     TCSKAIKIFMGRTELKLEDKHRVVIQRDEGHHVAYTTREVGQYLVVESSTGIIVIWDK
                     RTTVFIKLAPSYKGTVCGLCGNFDHRSNNDFTTRDHMVVSSELDFGNSWKEAPTCPDV
                     STNPEPCSLNPHRRSWAEKQCSILKSSVFSICHSKVDPKPFYEACVHDSCSCDTGGDC
                     ECFCSAVASYAQECTKEGACVFWRTPDLCPIFCDYYNPPHECEWHYEPCGNRSFETCR
                     TINGIHSNISVSYLEGCYPRCPKDRPIYEEDLKKCVTADKCGCYVEDTHYPPGASVPT
                     EETCKSCVCTNSSQVVCRPEEGKILNQTQDGAFCYWEICGPNGTVEKHFNICSITTRP
                     STLTTFTTITLPTTPTSFTTTTTTTTPTSSTVLSTTPKLCCLWSDWINEDHPSSGSDD
                     GDREPFDGVCGAPEDIECRSVKDPHLSLEQHGQKVQCDVSVGFICKNEDQFGNGPFGL
                     CYDYKIRVNCCWPMDKCITTPSPPTTTPSPPPTTTTTLPPTTTPSPPTTTTTTPPPTT
                     TPSPPITTTTTPLPTTTPSPPISTTTTPPPTTTPSPPTTTPSPPTTTPSPPTTTTTTP
                     PPTTTPSPPMTTPITPPASTTTLPPTTTPSPPTTTTTTPPPTTTPSPPTTTPITPPTS
                     TTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTPSPPTITTTTPPPTTTPSPPTTTTT
                     TPPPTTTPSPPTTTPITPPTSTTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPP
                     ITTTTTPPPTTTPSSPITTTPSPPTTTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPT
                     TMTTPSPTTTPSPPTTTMTTLPPTTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVP
                     LCNWTGWLDSGKPNFHKPGGDTELIGDVCGPGWAANISCRATMYPDVPIGQLGQTVVC
                     DVSVGLICKNEDQKPGGVIPMAFCLNYEINVQCCECVTQPTTMTTTTTENPTPPTTTP
                     ITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT
                     PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI
                     TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP
                     TGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPIT
                     TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPT
                     GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT
                     TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTG
                     TQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTT
                     TTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGT
                     QTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTT
                     TVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ
                     TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTT
                     VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT
                     PTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTV
                     TPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTP
                     TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT
                     PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT
                     IDESCQCQDTVCGLPTGTSRRARRSPRHLGSG"
     misc_feature    133. .1194
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     misc_feature    1195. .2295
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     misc_feature    2602. .3687
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     repeat_region   4228. .5268
                     /note="repetitive Thr/Ser/Pro-rich subdomain"
     misc_feature    13465. .14634
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
BASE COUNT     4107 a   6921 c   2862 g   1830 t
```

ORIGIN

Query Match 7.0%; Score 198; DB 11; Length 15720;
Best Local Similarity 46.4%; Pred. No. 5.5e-24;
Matches 908; Conservative 0; Mismatches 1025; Indels 25; Gaps 7;

```
Qy    244 CGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCTGCC 303
Db   5628 CAACTACGAGATCAACGTTCAGTGCTGTGAGTGTGTCACCCAACCCACCACCATGACAAC 5687

Qy    304 AGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCACCCA 363
Db   5688 CACCACCACAGAGAACCCAACTCCGCCAACCACGACACCCATCACCACCACCACTACGGT 5747

Qy    364 GGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAACCAC 423
Db   5748 GACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCAC 5807

Qy    424 CCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCAGGC 483
Db   5808 CACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCAT 5867

Qy    484 CCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCACCCA 543
Db   5868 CACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCAC 5927

Qy    544 GGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGCCCC 603
Db   5928 GACACCCATCACCACC---ACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACA 5984

Qy    604 AACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTACCCC 663
Db   5985 GACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCAC 6044

Qy    664 GGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGCACC 723
Db   6045 CGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCC 6104

Qy    724 CACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAG---- 780
Db   6105 AACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGAC 6164

Qy    780 -CTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAG 838
Db   6165 CCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCAC 6224

Qy    839 TGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTAC 898
Db   6225 TACGGTGACCCCAACCCCAACACCCACCGGCACACAGA---CCCCAACCACGACACCCAT 6281

Qy    899 GTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCC 958
Db   6282 CACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCAC' 6341

Qy    959 CCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGA 1018
Db   6342 GACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGAC 6401

Qy   1019 TGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTAT 1078
Db   6402 CCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGG 6461

Qy   1079 TACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGT 1138
Db   6462 CACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAAC 6521

Qy   1139 TTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAG 1198
Db   6522 ACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCC 6581

Qy   1199 AGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCC 1258
```

```
Db   6582 AACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTAC 6641

Qy   1259 ATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTC 1318
Db   6642 GGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCAC 6701

Qy   1319 CCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGA 1378
Db   6702 CACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACC--CCAACCACGACA 6759

Qy   1379 CCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACC 1438
Db   6760 CCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCA 6819

Qy   1439 GCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCG 1498
Db   6820 ACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACA 6879

Qy   1499 ACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCC 1558
Db   6880 CAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCC 6939

Qy   1559 TCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCC------AGTCCCACCT 1612
Db   6940 ACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACC 6999

Qy   1613 ACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCT 1672
Db   7000 CCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTG 7059

Qy   1673 GCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   7060 ACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACC 7119

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db   7120 ACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATC 7179

Qy   1793 GCTGCCCCCACCGCTG---CCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCA 1849
Db   7180 ACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACG 7239

Qy   1850 ACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACC 1909
Db   7240 ACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACC 7299

Qy   1910 ACAGTCACTGTACCACCCACTGCTGCCCCTAC---TACCGCAGCACCTGCTCCCAACACC 1966
Db   7300 CCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGC 7359

Qy   1967 ACAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACA 2026
Db   7360 ACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACA 7419

Qy   2027 GTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCC 2086
Db   7420 CCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCA 7479

Qy   2087 AACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGAT 2146
Db   7480 ACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACG 7539

Qy   2147 GACATCGACCCCCCTCTCCCCAACGACCCCATCAACCC 2184
Db   7540 GTGACCCCAACCCCAACACCCACCGGCACACAGACCCC 7577
```

```
RESULT   4
HUMMUC2X
LOCUS       HUMMUC2X    15720 bp    mRNA            PRI       09-JAN-1998
DEFINITION  Homo sapiens intestinal mucin (MUC2) mRNA, complete cds.
ACCESSION   L21998
NID         g454153
```

```
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 15720)
  AUTHORS   Gum,J.R.Jr. ., Hicks,J.W., Toribara,N.W., Kim,Y.S. and Siddiki,B.
  TITLE     Molecular cloning of human intestinal mucin (MUC2) cDNA:
            Identification of the amino terminal and overall sequence
            similarity to pre-pro-von Willebrand factor
  JOURNAL   J. Biol. Chem. 269, 2440-2446 (1994)
  MEDLINE   94132002
REFERENCE   2  (bases 1 to 15720)
  AUTHORS   Eckhardt,A.E., Timpte,C.S., DeLuca,A.W. and Hill,R.L.
  TITLE     The complete cDNA sequence and structural polymorphism of the
            polypeptide chain of porcine submaxillary mucin
  JOURNAL   J. Biol. Chem. 272 (52), 33204-33210 (1997)
  MEDLINE   98070526
FEATURES             Location/Qualifiers
     source          1. .15720
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="intestine"
     CDS             28. .15567
                     /codon_start=1
                     /product="mucin"
                     /db_xref="PID:g454154"
                     /translation="MGLPLARLAAVCLALSLAGGSELQTEGRTRYHGRNVCSTWGNFH
                     YKTFDGDVFRFPGLCDYNFASDCRGSYKEFAVHLKRGPGQAEAPAGVESILLTIKDDT
                     IYLTRHLAVLNGAVVSTPHYSPGLLIEKSDAYTKVYSRAGLTLMWNREDALMLELDTK
                     FRNHTCGLCGDYNGLQSYSEFLSDGVLFSPLEFGNMQKINQPDVVCEDPEEEVAPASC
                     SEHRAECERLLTAEAFADCQDLVPLEPYLRACQQDRCRCPGGDTCVCSTVAEFSRQCS
                     HAGGRPGNWRTATLCPKTCPGNLVYLESGSPCMDTCSHLEVSSLCEEHRMDGCFCPEG
                     TVYDDIGDSGCVPVSQCHCRLHGHLYTPGQEITNDCEQCVCNAGRWVCKDLPCPGTCA
                     LEGGSHITTFDGKTYTFHGDCYYVLAKGDHNDSYALLGELAPCGSTDKQTCLKTVVLL
                     ADKKKNAVVFKSDGSVLLNQLQVNLPHVTASFSVFRPSSYHIMVSMAIGVRLQVQLAP
                     VMQLFVTLDQASQGQVQGLCGNFNGLEGDDFKTASGLVEATGAGFANTWKAQSTCHDK
                     LDWLDDPCSLNIESANYAEHWCSLLKKTETPFGRCHSAVDPAEYYKRCKYDTCNCQNN
                     EDCLCAALSSYARACTAKGVMLWGWREHVCNKDVGSCPNSQVFLYNLTTCQQTCRSLS
                     EADSHCLEGFAPVDGCGCPDHTFLDEKGRCVPLAKCSCYHRGLYLEAGDVVVRQEERC
                     VCRDGRLHCRQIRLIGQSCTAPKIHMDCSNLTALATSKPRALSCQTLAAGYYHTECVS
                     GCVCPDGLMDDGRGGCVVEKECPCVHNNDLYSSGAKIKVDCNTCTCKRGRWVCTQAVC
                     HGTCSIYGSGHYITFDGKYYDFDGHCSYVAVQDYCGQNSSLGSFSIITENVPCGTTGV
                     TCSKAIKIFMGRTELKLEDKHRVVIQRDEGHHVAYTTREVGQYLVVESSTGIIVIWDK
                     RTTVFIKLAPSYKGTVCGLCGNFDHRSNNDFTTRDHMVVSSELDFGNSWKEAPTCPDV
                     STNPEPCSLNPHRRSWAEKQCSILKSSVFSICHSKVDPKPFYEACVHDSCSCDTGGDC
                     ECFCSAVASYAQECTKEGACVFWRTPDLCPIFCDYYNPPHECEWHYEPCGNRSFETCR
                     TINGIHSNISVSYLEGCYPRCPKDRPIYEEDLKKCVTADKCGCYVEDTHYPPGASVPT
                     EETCKSCVCTNSSQVVCRPEEGKILNQTQDGAFCYWEICGPNGTVEKHFNICSITTRP
                     STLTTFTTITLPTTPTSFTTTTTTTTPTSSTVLSTTPKLCCLWSDWINEDHPSSGSDD
                     GDREPFDGVCGAPEDIECRSVKDPHLSLEQHGQKVQCDVSVGFICKNEDQFGNGPFGL
                     CYDYKIRVNCCWPMDKCITTPSPPTTTPSPPPTTTTLPPTTTPSPPTTTTTTPPPTT
                     TPSPPITTTTTPLPTTTPSPPISTTTPPPTTTPSPPTTTPSPPTTTPSPPTTTTTTP
                     PPTTTPSPPMTTPITPPASTTTLPPTTTPSPPTTTTTTPPPTTTPSPPTTTPITPPTS
                     TTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPPTITTTPPPTTTPSPPTTTTT
                     TPPPTTTPSPPTTTPITPPTSTTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPP
                     ITTTTTPPPTTTPSSPITTTPSPPTTTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPT
                     TMTTPSPTTTPSPPTTTMTTLPPTTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVP
                     LCNWTGWLDSGKPNFHKPGGDTELIGDVCGPGWAANISCRATMYPDVPIGQLGQTVVC
                     DVSVGLICKNEDQKPGGVIPMAFCLNYEINVQCCECVTQPTTMTTTTTENPTPPTTTP
                     ITTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPT
                     PTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPI
                     TTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTP
                     TGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPIT
                     TTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPT
                     GTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITT
                     TTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTG
                     TQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTT
                     TTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGT
                     QTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTT
                     TVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQ
                     TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTT
                     VTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQT
                     PTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTV
                     TPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTP
                     TTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTV
                     PTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTGPPTHTSTAPIAELTTSNPPPESST
                     PQTSRSTSSPLTESTTLLSTLPPAIEMTSTAPPSTPTAPTTTSGGHTLSPPPSTTTSP
                     PGTPTRGTTTGSSSAPTPSTVQTTTTSAWTPTPTPLSTPSIIRTTGLRPYPSSVLICC
                     VLNDTYYAPGEEVYNGTYGDTCYFVNCSLSCTLEFYNWSCPSTPSPTPTPSKSTPTPS
                     KPSSTPSKPTPGTKPPECPDFDPPRQENETWWLCDCFMATCKYNNTVEIVKVECEPPP
                     MPTCSNGLQPVRVEDPDGCCWHWECDCYCTGWGDPHYVTFDGLYYSYQGNCTYVLVEE
                     ISPSVDNFGVYIDNYHCDPNDKVSCPRTLIVRHETQEVLIKTVHMMPMQVQVQVNRQA
                     VALPYKKYGLEVYQSGINYVVDIPELGVLVSYNGLSFSVRLPYHRFGNNTKGQCGTCT
                     NTTSDDCILPSGEIVSNCEAAADQWLVNDPSKPHCPHSSSTTKRPAVTVPGGGKTTPH
                     KDCTPSPLCQLIKDSLFAQCHALVPPQHYYDACVFDSCFMPGSSLECASLQAYAALCA
                     QQNICLDWRNHTHGACLVECPSHREYQACGPAEEPTCKSSSSQQNNTVLVEGCFCPEG
                     TMNYAPGFDVCVKTCGCVGPDNVPREFGEHFEFDCKNCVCLEGGSGIICQPKRCSQKP
                     VTHCVEDGTYLATEVNPADTCCNITVCKCNTSLCKEKPSVCPLGFEVKSKMVPGRCCP
                     FYWCESKGVCVHGNAEYQPGSPVYSSKCQDCVCTDKVDNNTLLNVIACTHVPCNTSCS
                     PGFELMEAPGECCKKCEQTHCIIKRPDNQHVILKPGDFKSDPKNNCTFFSCVKIHNQL
                     ISSVSNITCPNFDASICIPGSITFMPNGCCKTCTPRNETRVPCSTVPVTTEVSYAGCT
                     KTVLMNHCSGSCGTFVMYSAKAQALDHSCSCCKEEKTSQREVVLSCPNGGSLTHTYTH
                     IESCQCQDTVCGLPTGTSRRARRSPRHLGSG"
     misc_feature    133. .1194
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     misc_feature    1195. .2295
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     misc_feature    2602. .3687
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     repeat_region   4228. .5268
                     /note="repetitive Thr/Ser/Pro-rich subdomain"
     misc_feature    13465. .14634
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
BASE COUNT     4107 a   6921 c   2862 g   1830 t
ORIGIN
```

Query Match 7.0%; Score 198; DB 41; Length 15720;
Best Local Similarity 46.4%; Pred. No. 5.5e-24;
Matches 908; Conservative 0; Mismatches 1025; Indels 25; Gaps 7;

```
Qy    244 CGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCTGCC 303
          | | | | || ||  || ||||   ||||     | | |   | ||     ||      |
Db   5628 CAACTACGAGATCAACGTTCAGTGCTGTGAGTGTGTCACCCAACCCACCACCATGACAAC 5687

Qy    304 AGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCACCCA 363
              |  |||   |  |||| | | |  |  | ||| |     || | || || ||
Db   5688 CACCACCACAGAGAACCCAACTCCGCCAACCACGACACCCATCACCACCACCACTACGGT 5747
```

```
Qy    364 GGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAACCAC 423
Db   5748 GACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCAC 5807

Qy    424 CCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCAGGC 483
Db   5808 CACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCAT 5867

Qy    484 CCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCACCCA 543
Db   5868 CACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCAC 5927

Qy    544 GGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGCCCC 603
Db   5928 GACACCCATCACCACC---ACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACA 5984

Qy    604 AACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTACCCC 663
Db   5985 GACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCAC 6044

Qy    664 GGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGCACC 723
Db   6045 CGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCC 6104

Qy    724 CACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAG---- 780
Db   6105 AACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGAC 6164

Qy    780 -CTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAG 838
Db   6165 CCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCAC 6224

Qy    839 TGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTAC 898
Db   6225 TACGGTGACCCCAACCCCAACACCCACCGGCACACAGA---CCCCAACCACGACACCCAT 6281

Qy    899 GTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCC 958
Db   6282 CACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCAC 6341

Qy    959 CCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGA 1018
Db   6342 GACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGAC 6401

Qy   1019 TGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTAT 1078
Db   6402 CCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGG 6461

Qy   1079 TACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGT 1138
Db   6462 CACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAAC 6521

Qy   1139 TTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAG 1198
Db   6522 ACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCC 6581

Qy   1199 AGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCC 1258
Db   6582 AACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTAC 6641

Qy   1259 ATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTC 1318
Db   6642 GGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCAC 6701

Qy   1319 CCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGA 1378
Db   6702 CACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACC--CCAACCACGACA 6759

Qy   1379 CCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACC 1438
Db   6760 CCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCA 6819

Qy   1439 GCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCG 1498
Db   6820 ACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACA 6879

Qy   1499 ACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCC 1558
Db   6880 CAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCC 6939

Qy   1559 TCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCC------AGTCCCACCT 1612
Db   6940 ACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACC 6999

Qy   1613 ACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCT 1672
Db   7000 CCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTG 7059

Qy   1673 GCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   7060 ACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACC 7119

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db   7120 ACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATC 7179

Qy   1793 GCTGCCCCCACCGCTG---CCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCA 1849
Db   7180 ACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACG 7239

Qy   1850 ACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACC 1909
Db   7240 ACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACC 7299

Qy   1910 ACAGTCACTGTACCACCCACTGCTGCCCCTAC---TACCGCAGCACCTGCTCCCAACACC 1966
Db   7300 CCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGC 7359

Qy   1967 ACAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACA 2026
Db   7360 ACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACA 7419

Qy   2027 GTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCC 2086
Db   7420 CCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCA 7479

Qy   2087 AACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGAT 2146
Db   7480 ACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACG 7539

Qy   2147 GACATCGACCCCCCTCTCCCCAACGACCCCATCAACCC 2184
Db   7540 GTGACCCCAACCCCAACACCCACCGGCACACAGACCCC 7577
```

```
RESULT   5
AC002042/c
LOCUS       AC002042   185994 bp    DNA             HTG       22-APR-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 16p11.2 BAC
            clone CIT987SK-A-180G2; HTGS phase 1, 5 unordered pieces.
ACCESSION   AC002042
NID         g3075381
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185994)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Phillips,C., Brandon,R.C.,
            Fuhrmann,J., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Human chromosome 16p13 BAC clone CIT987SK-A-180G2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185994)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (29-APR-1997) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Apr 22, 1998 this sequence version replaced gi:3068565.
            ***                                                        ***
            *** WARNING: Phase 1 High Throughput Genome Sequence ***
            ***                                                        ***
            * This sequence is unfinished. It consists of 5 contigs for
            * which the order is not known; their order in this record is
            * arbitrary. In some cases, the exact lengths of the gaps
            * between the contigs are also unknown; these gaps are  presented
            * as runs of N as a convenience only. When sequencing is complete,
            * the sequence data presented in this record will be replaced
            *by a single finished sequence with the same accession number.
            *        1     2198: contig of 2198 bp in length
            *     2199     2248: gap of unknown length
            *     2249    46288: contig of 44040 bp in length
            *    46289    46338: gap of unknown length
            *    46339   110919: contig of 64581 bp in length
            *   110920   110969: gap of unknown length
            *   110970   145225: contig of 34256 bp in length
            *   145226   145275: gap of unknown length
            *   145276   185994: contig of 40719 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185994
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="A-180G2"
BASE COUNT    51275 a  43334 c  43047 g  48137 t    201 others
ORIGIN


  Query Match             5.9%;  Score 166.4;  DB 18;  Length 185994;
  Best Local Similarity   55.6%;  Pred. No. 7.7e-19;
  Matches  367;  Conservative    0;  Mismatches  281;  Indels   12;  Gaps    2;

Qy   1485 CGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCC 1544
          | ||| || |  | ||||||||| |||| || |   |  | || |  | || ||| |
Db 112122 CTACTCCTCCTGCCACAGCTGCTGCAACTGCCGATGCTGCTACTCCTGCCACAGCTACTG 112063

Qy   1545 CAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAG 1604
          ||||||| |   |  | ||       || ||||| ||  |  | |||||||| |   | |
Db 112062 CAACTGCTGATGCTGCTAC------TCCTCCTGCCACAGCTACTGCAACTGCTGATGCTG 112009

Qy   1605 TCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTA 1664
            | ||| |  || ||  |  | |  | ||  | ||  || |  |||| ||  || | |
Db 112008 CTACTCCTCCTGCCACAGCTATTGCAACTGCTGCTACTCCTCCTGCCACAGCTACTGCAA 111949

Qy   1665 CTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCG 1724
          || ||||| | |||    || || | ||||  || |   |  || || || ||||  ||
Db 111948 CTGCTGCTACTCCTCCTGCCACAGCTACTGCAACTGCTGCTGCTGCTACTCCTCCTGCCA 111889

Qy   1725 CAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTG 1784
          ||||  || | ||| |||   ||| |  ||   ||| |  || || |  ||   ||||
Db 111888 CAGCTACTGCAACTACTGATGCTGCTACTCCTGCCACAGCTACTGCAACTGCTGATGCTG 111829

Qy   1785 CCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAA 1844
          |  |  |  ||||| |  |  ||||  |  | |  ||  ||||| || ||| | |||  |
Db 111828 CTACTCCTCCTGCCACAGCTACTGCAACTGCTGATGCTGCTACTCCTCCTGCCACAGCTA 111769

Qy   1845 TCCCAAC----:--TACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTG 1898
            ||||       | |||  ||  | ||  | || ||| |  | ||| | || ||  ||
Db 111768 CTGCAACTGCTGATGCTGCTACTCCTCCTGCCACAGCTACTGCAACTGCTGCTGCTGCTA 111709

Qy   1899 CCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTC 1958
          | ||     ||||||  |||| | |  ||  ||||||  || || || ||||  ||||
Db 111708 CTCCTCCTGCCACAGCTACTGCAACTGCCGATGCTGCTACTCCTGCCACAGCTACTGCAA 111649

Qy   1959 CCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCA 2018
          |        | |  ||| | ||  |||| ||  |  | ||| | |  ||  || | ||
Db 111648 CTGCTGATGCTGCTACTCCTCCTGCCACAGCTACTGCAACTGCTGATGCTGCTACTCCTC 111589

Qy   2019 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2078
            ||||||  |||| | |  |   ||||||  |  || | ||  |  |||| |   |
```

```
Db 111588 CTGCCACAGCTACTGCAACTGCTGATGCTGCTACTCCTCCTGCCACAGCTACTGCAACTG 111529

Qy   2079 CTGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2138
          |||   |    ||  |  ||||| | |  ||| ||||    |   | |||||  | ||||
Db 111528 CTGATGCTGCTACTCCTCCTGCCACAGCTACTGCAACTGCTGATGCTGCTACTCCTCCTG 111469


RESULT   6
I25041/c
LOCUS       I25041      10596 bp    DNA             PAT       21-AUG-1996
DEFINITION  Sequence 15 from patent US 5547856.
ACCESSION   I25041
NID         g1604911
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10596)
  AUTHORS   Godowski,P.J., Lokker,N.A. and Mark,M.R.
  TITLE     Hepatocyte growth factor variants
  JOURNAL   Patent: US 5547856-A 15 20-AUG-1996;
FEATURES             Location/Qualifiers
     source          1. .10596
                     /organism="unknown"
BASE COUNT     2627 a   2571 c   3023 g   2375 t
ORIGIN


  Query Match             5.6%;  Score 158.8;  DB 6;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 1.8e-17;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  | |
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || | .| |      | ||| |  | || || ||  ||||  | || | ||| ||| |
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
          | ||   |||| | ||  | ||  |  |||||||  | ||  |||||  || || |||||
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||   | ||  |  ||| | |     | ||| ||||| || || |  | ||||||
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
          || || ||| ||||      ||   | ||    |  | || ||| | ||| |  | | |
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
          |||||||  | ||  | ||  | ||  |  |  || |||||    |         |  ||
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
          |||  | |   |  | |||  |  || | ||  || | ||  | ||||||||     || |
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
            |||  ||  || || |||||||| || | ||  |  |  ||||      |  | |
Db   2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
          || |  | ||  | ||  | ||| || |  | || ||| |  ||     |  | |  ||
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
          |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||   | |
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
```

```
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
Db   2286 CCTCCTCCTGCTCC 2273


RESULT   7
I30503/c
LOCUS       I30503      10596 bp    DNA             PAT       13-DEC-1996
DEFINITION  Sequence 15 from patent US 5580963.
ACCESSION   I30503
NID         g1821294
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10596)
  AUTHORS   Godowski,P.J., Lokker,N.A. and Mark,M.R.
  TITLE     Single-chain hepatocyte growth factor variants
  JOURNAL   Patent: US 5580963-A 15 03-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .10596
                     /organism="unknown"
BASE COUNT     2627 a    2571 c    3023 g    2375 t
ORIGIN


  Query Match             5.6%;  Score 158.8;  DB 6;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 1.8e-17;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
Db   2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
Db   2286 CCTCCTCCTGCTCC 2273


RESULT   8
U02454/c
LOCUS       U02454      5452 bp    DNA   circular  SYN       29-MAR-1996
DEFINITION  Cloning vector pCMVEBNA, complete sequence.
ACCESSION   U02454
NID         g413820
KEYWORDS    .
SOURCE      Cloning vector pCMVEBNA.
  ORGANISM  Cloning vector pCMVEBNA
            artificial sequence; cloning vectors.
REFERENCE   1  (bases 1 to 5452)
  AUTHORS   Kitts,P.A.
  TITLE     CLONTECH Vectors On Disc version 1.3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5452)
  AUTHORS   Swirski,R.A., Van Den Berg,D., Murphy,A.J., Lambert,C.M.,
            Friedberg,E.C. and Schimke,R.T.
  TITLE     Improvements in the Epstein-Barr-based shuttle vector system for
            direct cloning in human tissue culture cells
  JOURNAL   Methods: A Companion to Methods in Enzymology 4, 133-142 (1992)
REFERENCE   3  (bases 1 to 5452)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT     This vector can be obtained from CLONTECH Laboratories, Inc., 1020
            East Meadow Circle, Palo Alto, CA 94303, USA. To place an order
            call (415) 424-8222 or (800) 662-2566, extension 1. International
            customers, please contact your local distributor.  For technical
            information, call (415) 424- 8222 or (800) 662-2566, extension 3.
            This sequence has been compiled from information in the sequence
            databases, published literature and other sources, together with
            partial sequences obtained by CLONTECH; this vector has not been
            completely sequenced. If you suspect there is an error in this
            sequence, please contact CLONTECH's Technical Service Department at
            (415) 424-8222 or (800) 662-2566, extension 3 or E-mail
            TECH@CLONTECH.COM.
FEATURES             Location/Qualifiers
     source          1. .5452
                     /organism="Cloning vector pCMVEBNA"
                     /db_xref="taxon:31797"
BASE COUNT     1362 a    1246 c    1736 g    1108 t
ORIGIN


  Query Match             5.6%;  Score 158.8;  DB 15;  Length 5452;
  Best Local Similarity   53.3%;  Pred. No. 1.9e-17;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
Db   1780 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 1721

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
Db   1720 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 1661

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db   1660 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 1601

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   1600 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 1541
```

```
Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
          || || ||| ||||      ||   | ||    |  | || ||| | ||| |  || |
Db   1540 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 1481

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
          |||||||  | ||  | ||  | ||  |  | || ||||     |          |  ||
Db   1480 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 1421

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCCAACACCACA 1912
          ||| | |  |  | ||| |  || | ||  || | ||  | ||||||||     || |
Db   1420 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 1361

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
             |||  ||  || || |||||||| || | ||  |  |  ||||     |  | |
Db   1360 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 1301

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
          || |  | ||  | ||  | ||| || |  | || ||| |  ||     |  | |   ||
Db   1300 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 1241

Qy   2033 GTACCACCCACTGCTGCCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
          |  || ||  ||||| |  || || | || || ||| ||    | |||||  ||    | |
Db   1240 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 1181

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
          |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db   1180 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 1121

Qy   2153 GACCCCCCTCTCCC 2166
             || |||   ||
Db   1120 CCTCCTCCTGCTCC 1107
```

```
RESULT   9
U02455/c
LOCUS       U02455      10850 bp    DNA   circular  SYN       29-MAR-1996
DEFINITION  Cloning vector rpDR2, complete sequence.
ACCESSION   U02455
NID         g413821
KEYWORDS    .
SOURCE      Cloning vector rpDR2.
  ORGANISM  Cloning vector rpDR2
            artificial sequence; cloning vectors.
REFERENCE   1  (bases 1 to 10850)
  AUTHORS   Kitts,P.A.
  TITLE     CLONTECH Vectors On Disc version 1.3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 10850)
  AUTHORS   Murphy,A.J., Kung,A.L., Swirski,R.A. and Schimke,R.T.
  TITLE     cDNA expression cloning in human cells using the plambdaDR2
            episomal vector system
  JOURNAL   Methods: A Companion to Methods in Enzymology 4, 111-131 (1992)
REFERENCE   3  (bases 1 to 10850)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT     Lambda DR2 can be obtained from CLONTECH Laboratories, Inc., 1020
            East Meadow Circle, Palo Alto, CA 94303, USA. To place an order
            call (415) 424-8222 or (800) 662-2566, extension 1. International
            customers, please contact your local distributor.  For technical
            information, call (415) 424- 8222 or (800) 662-2566, extension 3.
            This sequence was compiled by Andrew Murphy and revised at
            CLONTECH. If you suspect there is an error in this sequence, please
            contact CLONTECH's Technical Service Department at (415) 424-8222
            or (800) 662-2566, extension 3 or E-mail TECH@CLONTECH.COM.
FEATURES             Location/Qualifiers
     source          1. .10850
                     /organism="Cloning vector rpDR2"
                     /note="plasmid released from lambda DR2"
                     /db_xref="taxon:31858"
BASE COUNT     2723 a   2615 c   3053 g   2459 t
```

```
ORIGIN


  Query Match             5.6%;  Score 158.8;  DB 15;  Length 10850;
  Best Local Similarity   53.3%;  Pred. No. 1.8e-17;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db   7426 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 7367

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || |  | |     | ||| |  | || || ||  ||||  | || | ||| ||| |
Db   7366 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 7307

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
          | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db   7306 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 7247

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||   | ||  |  ||| | |     | ||| ||||| || || |  | ||||||
Db   7246 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 7187

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
          || || ||| ||||      ||   | ||    |  | || ||| | ||| |  || |
Db   7186 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 7127

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
          |||||||  | ||  | ||  | ||  |  | || ||||     |          |  ||
Db   7126 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 7067

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
          ||| | |  |  | ||| |  || | ||  || | ||  | ||||||||     || |
Db   7066 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 7007

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
             |||  ||  || |||||||| || | ||  |  |  ||||     |  | |
Db   7006 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 6947

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
          || |  | ||  | ||  | ||| || |  | || ||| |  ||     |  | |   ||
Db   6946 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 6887

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
          |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||   | |
Db   6886 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 6827

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
          |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db   6826 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 6767

Qy   2153 GACCCCCCTCTCCC 2166
             || |||   ||
Db   6766 CCTCCTCCTGCTCC 6753
```

```
RESULT  10
XXU02428/c
LOCUS       XXU02428    10737 bp    DNA   circular  SYN       29-JAN-1997
DEFINITION  Cloning vector pDR2, complete sequence.
ACCESSION   U02428
NID         g413794
KEYWORDS    .
SOURCE      Cloning vector pDR2.
  ORGANISM  Cloning vector pDR2
            artificial sequence; cloning vectors.
REFERENCE   1  (bases 1 to 10737)
  AUTHORS   Murphy,A.J., Kung,A.L., Swirski,R.A. and Schimke,R.T.
  TITLE     cDNA expression cloning in human cells using the plambdaDR2
            episomal vector system
  JOURNAL   Methods: A Companion to Methods in Enzymology 4, 111-131 (1992)
REFERENCE   2  (bases 1 to 10737)
  AUTHORS   Kitts,P.A.
```

```
  TITLE     CLONTECH Vectors On Disc version 1.3
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 10737)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT     This vector can be obtained from CLONTECH Laboratories, Inc., 1020
            East Meadow Circle, Palo Alto, CA 94303, USA. To place an order
            call (415) 424-8222 or (800) 662-2566, extension 1. International
            customers, please contact your local distributor.  For technical
            information, call (415) 424- 8222 or (800) 662-2566, extension 3.
            This sequence has been compiled from information in the sequence
            databases, published literature and other sources, together with
            partial sequences obtained by CLONTECH; this vector has not been
            completely sequenced. If you suspect there is an error in this
            sequence, please contact CLONTECH's Technical Service Department at
            (415) 424-8222 or (800) 662-2566, extension 3 or E-mail
            TECH@CLONTECH.COM.'.
FEATURES              Location/Qualifiers
     source           1. .10737
                      /organism="Cloning vector pDR2"
                      /db_xref="taxon:31805"
BASE COUNT     2686 a    2596 c    3033 g    2422 t
ORIGIN


  Query Match             5.6%;  Score 158.8;  DB 15;  Length 10737;
  Best Local Similarity   53.3%;  Pred. No. 1.8e-17;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
Db   7313 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 7254

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
Db   7253 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 7194

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db   7193 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 7134

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   7133 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 7074

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db   7073 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 7014

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
Db   7013 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 6954

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
Db   6953 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 6894

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
Db   6893 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 6834

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
Db   6833 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 6774

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
Db   6773 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 6714

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
Db   6713 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 6654

Qy   2153 GACCCCCCTCTCCC 2166
Db   6653 CCTCCTCCTGCTCC 6640


RESULT  11
EBV/c
LOCUS       EBV        172281 bp    DNA   circular  VRL       17-AUG-1998
DEFINITION  Epstein-Barr virus (EBV) genome. The complete sequence [1-10] was
            determined from DNA from B95-8 cells cloned by Arrand et al [11].
            B95-8 is a productive marmoset lymphoblastoid cell line
            immortalized with human EBV from a mononucleosis patient.
ACCESSION   V01555 J02070 K01729 K01730 V01554 X00498 X00499 X00784
NID         g59074
KEYWORDS    DNA polymerase; EBNA; genome; ribonucleotide reductase; tandem
            repeat; terminal repeat.
SOURCE      Epstein-Barr virus.
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 172281)
  AUTHORS   Arrand,J.R., Rymo,L., Walsh,J.E., Bjorck,E., Lindahl,T. and
            Griffin,B.E.
  TITLE     Molecular cloning of the complete Epstein-Barr virus genome as a
            set of overlapping restriction endonuclease fragments
  JOURNAL   Nucleic Acids Res. 9 (13), 2999-3014 (1981)
  MEDLINE   82014887
REFERENCE   2  (bases 1 to 172281)
  AUTHORS   Kozak,M.
  TITLE     Possible role of flanking nucleotides in recognition of the AUG
            initiator codon by eukaryotic ribosomes
  JOURNAL   Nucleic Acids Res. 9 (20), 5233-5262 (1981)
  MEDLINE   82059504
REFERENCE   3  (bases 1 to 172281)
  AUTHORS   Deininger,P.L., Bankier,A., Farrell,P., Baer,R. and Barrell,B.
  TITLE     Sequence analysis and in vitro transcription of portions of the
            Epstein-Barr virus genome
  JOURNAL   J. Cell. Biochem. 19 (3), 267-274 (1982)
  MEDLINE   83109311
REFERENCE   4  (bases 1 to 172281)
  AUTHORS   Farrell,P.J., Deininger,P.L., Bankier,A. and Barrell,B.
  TITLE     Homologous upstream sequences near Epstein-Barr virus promoters
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 80 (6), 1565-1569 (1983)
  MEDLINE   83169725
REFERENCE   5  (bases 45644 to 52450)
  AUTHORS   Jeang,K.T. and Hayward,S.D.
  TITLE     Organization of the Epstein-Barr virus DNA molecule. III. Location
            of the P3HR-1 deletion junction and characterization of the NotI
            repeat units that form part of the template for an abundant
            12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript
  JOURNAL   J. Virol. 48 (1), 135-148 (1983)
  MEDLINE   83294686
REFERENCE   6  (bases 159853 to 172281)
  AUTHORS   Bankier,A.T., Deininger,P.L., Satchwell,S.C., Baer,R., Farrell,P.J.
            and Barrell,B.G.
  TITLE     DNA sequence analysis of the EcoRI Dhet fragment of B95-8
            Epstein-Barr virus containing the terminal repeat sequences
  JOURNAL   Mol. Biol. Med. 1 (4), 425-445 (1983)
  MEDLINE   85060428
REFERENCE   7  (bases 1 to 172281)
  AUTHORS   Farrell,P.J., Bankier,A.T., Seguin,C., Deininger,P.L. and
            Barrell,B.G.
  TITLE     Latent and lytic cycle promoters of Epstein-Barr virus
  JOURNAL   EMBO J. 2, 1331-1338 (1983)
REFERENCE   8  (bases 142687 to 159853)
  AUTHORS   Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G.
  TITLE     Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8
            Epstein-Barr virus
  JOURNAL   Mol. Biol. Med. 1 (1), 21-45 (1983)
  MEDLINE   85035713
REFERENCE   9  (bases 112620 to 125316)
```

AUTHORS Seguin,C., Farrell,P.J. and Barrell,B.G.
TITLE DNA sequence and transcription of the BamHI fragment B region of B95-8 Epstein-Barr virus
JOURNAL Mol. Biol. Med. 1 (3), 369-392 (1983)
MEDLINE 85060424
REFERENCE 10 (bases 45415 to 52824)
AUTHORS Jones,M.D., Foster,L., Sheedy,T. and Griffin,B.E.
TITLE The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (P3HR-1) of the virus
JOURNAL EMBO J. 3 (4), 813-821 (1984)
MEDLINE 84207939
REFERENCE 11 (bases 87650 to 92703)
AUTHORS Biggin,M., Farrell,P.J. and Barrell,B.G.
TITLE Transcription and DNA sequence of the BamHI L fragment of B95-8 Epstein-Barr virus
JOURNAL EMBO J. 3 (5), 1083-1090 (1984)
MEDLINE 84236104
REFERENCE 12 (bases 7315 to 9312)
AUTHORS Yates,J., Warren,N., Reisman,D. and Sugden,B.
TITLE A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)
MEDLINE 84222045
REFERENCE 13 (bases 76089 to 79808)
AUTHORS Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B.
TITLE Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes
JOURNAL Nucleic Acids Res. 12 (12), 5087-5099 (1984)
MEDLINE 84247360
REFERENCE 14 (bases 1 to 172281)
AUTHORS Baer,R., Bankier,A.T., Biggin,M.D., Deininger,P.L., Fallell,P.J., Gibson,T.J., Hatfull,G., Hudson,G.S., Satchwell,S.C., Seguin,C., Tuffnell,P.S. and Barrell,B.G.
TITLE DNA sequence and expression of the B95-8 Epstein-Barr virus genome
JOURNAL Nature 310 (5974), 207-211 (1984)
MEDLINE 84270667
REFERENCE 15 (bases 1 to 172281)
AUTHORS Bodescot,M. and Perricaudet,M.
TITLE Clustered alternative splice sites in Epstein-Barr virus RNAs
JOURNAL Nucleic Acids Res. 15 (14), 5887 (1987)
MEDLINE 87289053
REFERENCE 16 (bases 1 to 172281)
AUTHORS Laux,G., Perricaudet,M. and Farrell,P.J.
TITLE A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome
JOURNAL EMBO J. 7 (3), 769-774 (1988)
MEDLINE 88283646
REFERENCE 17 (bases 1 to 172281)
AUTHORS Hatfull,G.F., Barrell,B.G., Quinn,J. and McGeoch,D.
JOURNAL Unpublished
REFERENCE 18 (bases 1 to 172281)
AUTHORS Farrell,P.J. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-1984) to the EMBL/GenBank/DDBJ databases
REFERENCE 19 (bases 1 to 172281)
AUTHORS Farrell,P.J.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1988) Farrell P., Ludwig Institute for Cancer Research, St. Mary's Hospital Medical School, Norfolk Place London W2 1PG
COMMENT CDS

Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In general the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been excluded and on the other hand some small frames have been included which might represent exons or genes because they occur in a logical combination with other features or because of some other experimental data. The reading frames are named according to the Bam H1 fragment in which they start. eg BALF3 is the third leftward frame starting in Bam H1 fragment A. BORF1 is the first rightward frame in Bam H1 fragment O. If there is an obvious TATA sequence followed by an in frame Met codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

SITEs of POLYA signals

This feature lists all occurences of the sequence AATAAA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog ATTAAA is only listed when it is found in a position close to the end of a major reading frame.

SITEs of DONOR and ACCEPT sequences

This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case.

Restriction enzyme SITEs.

Only the positions of the sites Bam HI (BAM) are listed.

RPT

This feature is used to define repetitive sequences.

SITE DEL

This feature defines deletions in B95-8 with respect to other strains such as RAJI and also to deletions in other strains such as P3HR1 and DAUDI with respect to B95-8.

SITE HPN

Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurences noted.

ORGRPL

Denotes the region that encompasses.an origin of replication (ori P).[13].

NUMBERING

The DNA sequence of B95-8 EBV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCAGTCTTT. To avoid renumbering the entire sequence, position 1 has benn moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC).

FEATURES Location/Qualifiers
source 1. .172281
/organism="Human herpesvirus 4"
/strain="B95-8"
/db_xref="taxon:10376"

```
mRNA            58. .272
                /note="exon 2 terminal protein RNA"
mRNA            360. .458
                /note="exon 3 terminal protein RNA"
misc_feature    complement(535)
                /note="polyA signal: AATAAA"
mRNA            540. .788
                /note="exon 4 terminal protein RNA"
mRNA            871. .951
                /note="exon 5 terminal protein RNA"
mRNA            1026. .1196
                /note="exon 6 terminal protein RNA"
promoter        complement(1192)
                /note="TATA: TATAAAT"
mRNA            1280. .1495
                /note="exon 7 terminal protein RNA"
promoter        complement(1383)
                /note="TATA: CATAAAA"
mRNA            1574. .1682
                /note="exon 8 terminal protein RNA"
promoter        1676
                /note="TATA: TATAAAG"
promoter        1691
                /note="TATA: TATTAAA BN-R1 late promoter before BNRF1,
                gives 4.1kb late RNA. Probably encodes non glycosylated
                140kd protein in membrane antigen. Also two latent RNAs
                spliced underneath this RNA, lengths 1.8 and 2.0kb (Hudson
                et al, 1985). The longer one encodes terminal protein."
CDS             1736. .5692
                /note="BNRF1 reading frame, 5 NXT/S"
                /codon_start=1
                /db_xref="PID:g59075"
                /db_xref="SWISS-PROT:P03179"
                /translation="MEERGRETQMPVARYGGPFIMVRLFGQDGEANIQEERLYELLSD
                PRSALGLDPGPLIAENLLLVALRGTNNDPRPQRQERARELALVGILLGNGEQGEHLGT
                ESALEASGNNYVYAYGPDWMARPSTWSAEIQQFLRLLGATYVLRVEMGRQFGFEVHRS
                RPSFRQFQAINHLVLFDNALRKYDSGQVAAGFQRALLVAGPETADTRPDLRKLNEWVF
                GGRAAGGRQLADELKIVSALRDTYSGHLVLQPTETLDTWKVLSRDTRTAHSLEHGFIH
                AAGTIQANCPQLFMRRQHPGLFPFVNAIASSLGWYYQTATGPGADARAAARRQQAFQT
                RAAAECHAKSGVPVVAGFYRTINATLKGGEGLQPTMFNGELGAIKHQALDTVRYDYGH
                YLIMLGPFQPWSGLTAPPCPYAESSWAQAAVQTALELFSALYPAPCISGYARPPGPSA
                VIEHLGSLVPKGGLLLFLSHLPDDVKDGLGEMGPARATGPGMQQFVSSYFLNPACSNV
                FITVRQRGEKINGRTVLQALGRACDMAGCQHYVLGSTVPLGGLNFVNDLASPVSTAEM
                MDDFSPFFTVEFPPIQEEGASSPVPLDVDESMDISPSYELPWLSLESCLTSILSHPTV
                GSKEHLVRHTDRVSGGRVAQQPGVGPLDLPLADYAFVAHSQVWTRPGGAPPLPYRTWD
                RMTEKLLVSAKPGGENVKVSGTVITLGEQGYKVSLDLREGTRLAMAEALLNAACAPIL
                DPEDVLLTLHLHLDPRRADNSAVMEAMTAASDYARGLGVKLTFGSASCPETGSSASNF
                MTVVASVSAPGEFSGPLITPVLQKTGSLLIAVRCGDGKIQGGSLFEQLFSDVATTPRA
                PEALSLKNLFRAVQQLVKSGIVLSGHDISDGGLVTCLVEMALAGQRGVTITMPVASDY
                LPEMFAEHPGLVFEVEERSVGEVLQTLRSMNMYPAVLGRVGEQGPDQMFEVQHGPETV
                LRQSLRLLLGTWSSFASEQYECLRPDRINRSMHVSDYGYNEALAVSPLTGKNLSPRRL
                VTEPDPRCQVAVLCAPGTRGHESLLAAFTNAGCLCRRVFFREVRDNTFLDKYVGLAIG
                GVHGARDSALAGRATVALINRFPALRDAILKFLNRPDTFSVALGELGVQVLAGLGAVG
                STDNPPAPGVEVNVQRSPLILAPNASGMFESRWLNISIPATTSSVMLRGLRGCVLPCW
                VQGSCLGLQFTNLGMPYVLQNAHQIACHFHSNGTDAWRFAMNYPRNPTEQGNIAGLCS
                RDGRHLALLCDPSLCTDFWQWEHIPPAFGHPTGCSPWTLMFQAAHLWSLRHGRPSE"
misc_feature    complement(1795)
                /note="polyA signal: AATAAA"
misc_feature    3955
                /note="BAM: Bam H1 Nhet/h"
misc_feature    3994
                /note="BAM: Bam H1 h/C"
mRNA            5408. .5856
                /note="exon 9 terminal protein RNA"
misc_feature    5841
                /note="polyA signal: AATAAA, end of 4.1kb late RNA and TP
                latent RNA."
misc_feature    5863
                /note="alternative end to TP cDNAs"
promoter        6097
                /note="TATA: TATAAGA"
misc_feature    6629. .6795
                /note="Pol III RNA EBER 1"
promoter        complement(6823)
                /note="TATA: CATAAAT"
misc_feature    6956. .7128
                /note="Pol III RNA EBER 2"
rep_origin      7315. .9312
                /note="origin of replication, ori P (Yates et al, 1984,
                1985)"
repeat_region   7421. .8042
                /note="21x30bp repeats, binding sites for EBNA-1 (site I,
                Rawlins et al, 1985). Tandem repeat part of oriP (Reisman
                et al, 1985). Also functions as a cell type specific
                enhancer (Reisman et al, 1985; Lupton and Levine, 1985)"
promoter        7738
                /note="TATA: TATAAAT"
promoter        7888
                /note="TATA: TATAAAT"
promoter        8573
                /note="TATA: CATAAAT"
misc_feature    complement(8680)
                /note="polyA signal: AATAAA"
misc_feature    complement(8755)
                /note="polyA signal: AATAAA"
misc_feature    8962
                /note="polyA signal: AATAAA"
misc_feature    9021. .9133
                /note="HPN: dyad symmetry, site II for EBNA-1 binding
                (Rawlins et al, 1985). Dyad symmetry part of oriP (Reisman
                et al, 1985)"
promoter        complement(9398)
                /note="TATA: TATAAAT"
promoter        9631
                /note="TATA: TATAAAT BC-R1 late promoter before BCRF1"
CDS             9675. .10187
                /note="BCRF1 reading frame"
                /codon_start=1
                /db_xref="PID:g59076"
                /db_xref="SWISS-PROT:P03180"
                /translation="MERRLVVTLQCLVLLYLAPECGGTDQCDNFPQMLRDLRDAFSRV
                KTFFQTKDEVDNLLLKESLLEDFKGYLGCQALSEMIQFYLEEVMPQAENQDPEAKDHV
                NSLGENLKTLRLRLRRCHRFLPCENKSKAVEQIKNAFNKLQEKGIYKAMSEFDIFINY
                IEAYMTIKAR"
promoter        10076
                /note="TATA: GATAAAA"
misc_feature    complement(10148)
                /note="polyA signal: AATAAA"
misc_feature    10173
                /note="polyA signal: ATTAAA"
misc_feature    10257
                /note="polyA signal: AATAAA, end of 0.8kb late RNA from
                BCR1 and end of 1.6 kb late RNA, start unknown"
misc_feature    complement(10277)
                /note="polyA signal: AATAAA"
promoter        complement(10975)
                /note="TATA: CATAAAT"
promoter        11305
                /note="TATA: TACAAAA; BCR2 promoter for highly spliced
                EBNA latent RNAs."
mRNA            11336. .11480
                /note="exon C1 of Bodescot et al (1986) RNAs"
promoter        11524
                /note="TATA: TATAATT"
misc_feature    complement(11587)
                /note="polyA signal: AATAAA"
promoter        complement(11606)
                /note="TATA: CATAAAT"
mRNA            11626. .11657
                /note="exon C2 of Bodescot et al (1986) RNAs"
promoter        11796
                /note="TATA: TATAAGT"
promoter        complement(11799)
                /note="TATA: TATAAAA"
repeat_region   12001. .15072
                /note="3072 repeat 1"
```

```
CDS             12541. .13689
                /note="BCRF2 3072 repeat, reading frame 1; Protein
                sequence is in conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25077"
                /db_xref="PID:g1334836"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    13215
                /note="BAM: BamH1 C/W"
promoter        14352
                /note="*TATA: TATAAAG BWR1 one of the promoters for highly
                spliced EBNA and LP RNAs (Sample et al, 1986; Speck et al,
                1986)"
mRNA            14384. .14410
                /note="*exon W0 of EBNA/LP RNAs"
mRNA            14554. .14619
                /note="*exon W1 (also W66) part of leader protein (LP)
                gene. LP is also called EBNA-5 (Dillner et al, 1986) and
                EBNA-4 (Rowe et al, 1987)."
mRNA            14559. .14619
                /note="*exon W1' (also W61) of EBNA/LP RNAs forms
                initiator met when fused to exon W0 or exon C2."
mRNA            14701. .14832
                /note="*exon W2 (also W132) part of LP gene"
repeat_region   15073. .18144
                /note="3072 repeat 2"
CDS             15613. .16761
                /note="BWRF1 reading frame 2; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25078"
                /db_xref="PID:g1334837"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    16287
                /note="BAM: BamH1 W/W"
promoter        17424
                /note="TATA: TATAAAG"
mRNA            17626. .17691
                /note="Exon W1"
mRNA            17773. .17904
                /note="Exon W2"
repeat_region   18145. .21216
                /note="3072 repeat 3"
CDS             18685. .19833
                /note="BWRF1 reading frame 3; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25079"
                /db_xref="PID:g1334838"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    19359
                /note="BAM: BamH1 W/W"
promoter        20496
                /note="TATA: TATAAAG"
mRNA            20698. .20763
                /note="Exon W1"
mRNA            20845. .20976
                /note="Exon W2"
repeat_region   21217. .24288
                /note="3072 repeat 4"
CDS             21757. .22905
                /note="BWRF1 reading frame 4; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25080"
                /db_xref="PID:g1334839"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    22431
                /note="BAM: BamH1 W/W"
promoter        23568
                /note="TATA: TATAAAG"
mRNA            23771. .23835
                /note="Exon W1"
mRNA            23917. .24048
                /note="Exon W2"
repeat_region   24289. .27360
                /note="3072 repeat 5"
CDS             24829. .25977
                /note="BWRF1 reading frame 5; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25081"
                /db_xref="PID:g1334840"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    25503
                /note="BAM: BamH1 W/W"
promoter        26640
                /note="TATA: TATAAAG"
mRNA            26842. .26907
                /note="Exon W1"
mRNA            26989. .27120
                /note="Exon W2"
repeat_region   27361. .30432
                /note="3072 repeat 6"
CDS             27901. .29049
                /note="BWRF1 reading frame 6; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25066"
                /db_xref="PID:g1334841"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    28575
                /note="BAM: BamH1 W/W"
promoter        29712
```

```
                /note="TATA: TATAAAG"
mRNA            29914. .29979
                /note="Exon W1"
mRNA            30061. .30192
                /note="Exon W2"
repeat_region   30433. .33504
                /note="3072 repeat 7"
CDS             30973. .32121
                /note="BWRF1 reading frame 7; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25067"
                /db_xref="PID:g1334842"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    31647
                /note="BAM: BamH1 W/W"
promoter        32784
                /note="TATA: TATAAAG"
mRNA            32986. .33051
                /note="Exon W1"
mRNA            33133. .33264
                /note="Exon W2"
repeat_region   33505. .36576
                /note="3072 repeat 8"
CDS             34045. .35193
                /note="BWRF1 reading frame 8; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25068"
                /db_xref="PID:g1334843"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    34719
                /note="BAM: BamH1 W/W"
promoter        35856
                /note="TATA: TATAAAG"
mRNA            36058. .36123
                /note="Exon W1"
mRNA            36205. .36336
                /note="Exon W2"
repeat_region   36577. .39648
                /note="3072 repeat 9"
CDS             37117. .38265
                /note="BWRF1 reading frame 9; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25069"
                /db_xref="PID:g1334844"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    37791
                /note="BAM: BamH1 W/W"
promoter        38928
                /note="TATA: TATAAAG"
mRNA            39130. .39195
                /note="Exon W1"
mRNA            39277. .39408
                /note="Exon W2"
repeat_region   39649. .42720
                /note="3072 repeat 10"
CDS             40189. .41337
                /note="BWRF1 reading frame 10; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25070"
                /db_xref="PID:g1334845"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    40863
                /note="BAM: BamH1 W/W"
promoter        42000
                /note="TATA: TATAAAG"
mRNA            42202. .42267
                /note="Exon W1"
mRNA            42349. .42480
                /note="Exon W2"
repeat_region   42721. .45792
                /note="3072 repeat 11"
CDS             43261. .44409
                /note="BWRF1 reading frame 11; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25071"
                /db_xref="PID:g1334846"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    43935
                /note="BAM: BamH1 W/W"
promoter        45072
                /note="TATA: TATAAAG"
mRNA            45274. .45339
                /note="Exon W1"
misc_feature    45415. .52824
                /note="DEL: DAUDI deletion (Jones et al, 1984)"
mRNA            45421. .45552
                /note="Exon W2"
misc_feature    45644. .52450
                /note="DEL: P3HR1 deletion (Jeang and Hayward, 1983)"
repeat_region   45793. .47643
                /note="3072 repeat 12"
CDS             46333. .47481
                /note="BWRF1 reading frame 12; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25072"
                /db_xref="PID:g1334847"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    47007
```

/note="BAM: BamH1 W/Y"
mRNA 47761. .47793
/note="Exon Y1 Bodescot et al, 1984"
promoter 47831
/note="TATA: TATAAGT"
mRNA 47878. .47999
/note="Exon Y2 Bodescot et al, 1984 and EBNA-1 (Speck and Strominger,1985), last common exon"
misc_feature complement(48023)
/note="polyA signal: AATAAA"
mRNA 48386. .48444
/note="exon Bodescot et al, 1984"
CDS 48386. .50032
/note="Coding exon for EBNA-2 (Sample et al,1986)"
/codon_start=1
/db_xref="PID:e25073"
CDS 48504. .49967
/note="BYRF1, encodes EBNA-2 (Dambaugh et al, 1984; Dillner et al, 1984)"
/codon_start=1
/db_xref="PID:e25074"
/db_xref="PID:g1632787"
/db_xref="SWISS-PROT:P12978"
/translation="MPTFYLALHGGQTYHLIVDTDSLGNPSLSVIPSNPYQEQLSDTP
LIPLTIFVGENTGVPPPLPPPPPPPPPPPPPPPPPPPPPPPSPPPPPPPPPPQR
RDAWTQEPSPLDRDPLGYDVGHGPLASAMRMLWMANYIVRQSRGDRGLILPQGPQTAP
QARLVQPHVPPLRPTAPTILSPLSQPRLTPPQPLMMPPRPTPPTPLPPATLTVPPRPT
RPTTLPPTPLLTVLQRPTELQPTPSPPRMHLPVLHVPDQSMHPLTHQSTPNDPDSPEP
RSPTVFYNIPPMPLPPSQLPPPAAPAQPPPGVINDQQLHHLPSGPPWWPPICDPPQPS
KTQGQSRGQSRGRGRGRGRGRGKGKSRDKQRKPGGPWRPEPNTSSPSMPELSPVLGLH
QGQGAGDSPTPGPSNAAPVCRNSHTATPNVSPIHEPESHNSPEAPILFPDDWYPPSID
PADLDESWDYIFETTESPSSDEDYVEGPSKRPRPSIQ"
repeat_region 48678. .48800
/note="14 x CCCCCACCA repeats"
misc_feature 48848
/note="BAM: BamH1 Y/H"
promoter 49350
/note="TATA: TATAACA"
promoter complement(49353)
/note="TATA: TATAAAA"
repeat_region 49525. .49578
/note="9 x GGGGCA repeats"
mRNA 49852. .50032
/note="exon (Bodescot et al 1984)"
misc_feature 50003
/note="polyA signal: AATAAA, end of Bodescot T1 RNA and EBNA-2 RNA (3.0kb latent RNA in IB4 cells)"
promoter complement(50156)
/note="TATA: TATAAGT"
misc_feature complement(50317)
/note="polyA signal: AATAAA, end of 2.5kb early RNA from 52817"
misc_feature complement(50578. .52557)
/note="BHLF1 early reading frame"
repeat_region 50578. .52115
/note="12 x 125bp repeats"
misc_feature 52654. .53697
/note="region homologous to Eco R1 C of Raji"
promoter complement(52817)
/note="TATA: GATAAAA promoter for 2.5kb early RNA containing BHLF1 (Jeang and Hayward, 1983; Freese et al,1983)"
promoter 53759
/note="TATA: TATTAAC likely promoter for class III and IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
misc_feature 53895
/note="DONOR: CGGGTAACT donor for splice to 54335 in class IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
misc_feature 54335
/note="ACCEPT: TTTTCTAG acceptor from 48444 in class I, 47999 in class II, and 53895 in class IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
misc_feature 54376. .54948
/note="BHRF1 reading frame, limited homolgy to bcl-2 gene. Early gene in B95-8 cells and part of restricted EA complex."
promoter 54591
/note="TATA: TATAACA"
promoter complement(54594)
/note="TATA: TATAAAT"
misc_feature 54853
/note="BAM: BamH1 H/F"
misc_feature complement(54929)
/note="polyA signal: AATAAA"
promoter complement(54977)
/note="TATA: TATAAAG"
misc_feature 55518
/note="polyA signal: AATAAA, 3' end of 2.5kb, 1.9kb, 1.7kb and 0.6kb early RNAs"
misc_feature complement(55982. .56935)
/note="BFLF2 reading frame, 4 NXT/S, homologous to RF 27 in VZV and HFRF2 in CMV"
misc_feature complement(55990)
/note="polyA signal: AATAAA, 3' end of 2.3kb and 1.1kb early RNAs from 58568 and 57081"
promoter complement(56132)
/note="TATA: TATAAAG"
CDS complement(56948. .58525)
/note="BFLF1 reading frame, 2 NXT/S homologous to RF 26 in VZV and HFRF1 in CMV"
/codon_start=1
/db_xref="PID:g1334849"
/db_xref="SWISS-PROT:P03184"
/translation="MAHKVTSANEPNPLTGKRLSSCPLTRSGVTEVAQIAGRTPKMED
FVPWTVDNLKSQFEAVGLLMAHSYLPANAEEGIAYPPLVHTYESLSPASTCRVCDLLD
TLVNHSDAPVAFFEDYALLCYYCLNAPRAWISSLITGMDFLHILIKYFPMAGGLDSLF
MPSRILAIDIQLHFYICRCFLPVSSSDMIRNANLGYYKLEFLKSILTGQSPANFCFKS
MWPRTTPTFLTLPGPRTCKDSQDVPGDVGRGLYTALCCHLPTRNRVQHPFLRAEKGGL
SPEITTKADYCGLLLGTWQGTDLLGGPGHHAIGLNAEYSGDELAELALAITRPEAGDH
SQGPCLLAPMFGLRHKNASRTICPLCESLGAHPDAKDTLDRFKSLILDSFGNNIKILD
RIVFLIKTQNTLLDVPCPRLRAWLQMCTPQDFHKHLFCDPLCAINHSITNPSVLFGQI
YPPSFQAFKAALAAGQNLEQGVCDSLITLVYIFKSTQVARVGKTILVDVTKELDVVLR
IHGLDLVQSYQTSQVYV"
promoter complement(57081)
/note="TATA: TATTTAA before BFLF2; BFL2 promoter gives 1.1kb early RNA"
promoter complement(58088)
/note="TATA: GATAAAA"
promoter complement(58568)
/note="TATA: TATTAAA before BFLF1, BFL1 promoter gives 2.3kb early RNA"
promoter 58832
/note="TATA: TATAAAA before BFRF1"
CDS 58891. .59901
/note="BFRF1 early reading frame, 1 NXT/S, homologous to HFLF4 in CMV"
/codon_start=1
/db_xref="PID:g1334850"
/db_xref="SWISS-PROT:P03185"
/translation="MASPEERLLDELNNVIVSFLCDSGSLEVERCSGAHVFSRGSSQP
LCTVKLRHGQIYHLEFVYKFLAFKLKNCNYPSSPVFVISNNGLATTLRCFLHEPSGLR
SGQSGPCLGLSTDVDLPKNSIIMLGQDDFIKFKSPLVFPAELDLLKSMVVCRAYITEH
RTTMQFLVFQAANAQKASRVMDMISDMSQQLSRSGQVEDTGARVTGGGGPRPGVTHSG
CLGDSHVRGRGGWDLDNFSEAETEDEASYAPWRDKDSWSESEAAPWKKELVRHPIRRH
RTRETRRMRGSHSRVEHVPPETRETVVGGAWRYSWRATPYLARVLAVTAVALLLMFLR
WT"
CDS 59808. .61583
/note="BFRF2 early reading frame, homologous to HFLF5 in CMV"
/codon_start=1
/db_xref="PID:e25058"
/db_xref="PID:g1632788"
/db_xref="SWISS-PROT:P14347"
/translation="MALFLARHTLSGTGAGCHGRGPAPDVSEVDLTLQALGERGFSRL
LDLGLACLDLSYVEMREFVVWGRPPASEAAVASTPGSLFRSHSSAYWLSEVERPGGLV
RWARSQTSPSSLTLAPHLGPSLLSLSVVTGGGCGAVAFCNAFFLAYFLVVRSVFPAFS

DRIAAWICDRSPFCENTRAVARGYRGLVKRFLAFVFERSSYDPPLLRQNSRPVERCFA
IKNYVPGLDSQSCVTVPSFSRWAQSHASELDPREIRDRVTPATAPSFVADHASALLAS
LQKKASDTPCGNPIQWMWYRLLVNSCLRSAHCLLPIPAVSEGGRKTGGGVGEELVGAG
GPCLSRDVFVAIVSRNVLSCLLNVPAAGPRAYKCFRSHASRPVSGPDYPPLAVFCMDC
GYCLNFGKQTGVGGRLNSFRPTLQFYPRDQKEKHVLTCHASGRVYCSNCGSAAVGCQR
LAEPPSARSGWRPRIRAVLPHNAAYELDRGSRLLDAIIPCLGPDRTCMRPVVLRGVTV
RQLLYLTLRTEARAVCSICQQRQAPEDARDEPHLFSSCLEVELPPGERCAGCRLYQTR
YGTPAAQAHPPGEAGGGFSRQSPAS"

promoter complement(61062)
/note="TATA: GATAAAA"

promoter 61344
/note="TATA: TATTTAA before BFRF3"

CDS 61507. .62037
/note="BFRF3 early reading frame"
/codon_start=1
/db_xref="PID:e25059"
/db_xref="PID:g1632789"
/db_xref="SWISS-PROT:P14348"
/translation="MARRLPKPTLQGRLEADFPDSPLLPKFQELNQNNLPNDVFREAQ
RSYLVFLTSQFCYEEYVQRTFGVPRRQRAIDKRQRASVAGAGAHAHLGGSSATPVQQA
QAAASAGTGALASSAPSTAVAQSATPSVSSSISSLRAATSGATAAASAAAAVDTGSGG
GGQPHDTAPRGARKKQ"

misc_feature complement(62068)
/note="polyA signal: AATAAA"

misc_feature 62069
/note="polyA signal: AATAAA, 3' end of 10, 6.5, 3.7, 3.4, 3.1, 2.5 and 0.8kb early RNAs"

CDS complement(62078. .71527)
/note="BPLF1 reading frame, 1 NXT/S, analogous to VZV RF22"
/codon_start=1
/db_xref="PID:g1334853"
/db_xref="SWISS-PROT:P03186"
/translation="MSNGDWGQSQRTRGTGPVRGIRTMDVNAPGGGSGGSALRILGTA
SCNQAHCKFGRFAGIQCVSNCVLYLVKSFLAGRPLTSRPELDEVLDEGARLDALMRQS
GILKGHEMAQLTDVPSSVVLRGGGRVHIYRSAEIFGLVLFPAQIANSAVVQSLAEVLH
GSYNGVAQFILYICDIYAGAIIIETDGSFYLFDPHCQKDAAPGTPAHVRVSTYAHDIL
QYVGAPGAQYTCVHLYFLPEAFETEDPRIFMLEHYGVYDFYEANGSGFDLVGPELVSS
DGEAAGTPGADSSPPVMLPFERRIIPYNLRPLPSRSFTSDSFPAARYSPAKTNSPPSS
PASAAPASAAPASAAPASAAPASAAPASAAPASAAPASAAPASSPPLFIPIPGLGHTP
GVPAPSTPPRASSGAAPQTPKRKKGLGKDSPHKKPTSGRRLPLSSTTDTEDDQLPRTH
VPPHRPPSAARLPPPVIPIPHQSPPASPTPHPAPVSTIAPSVTPSPRLPLQIPIPLPQ
AAPSNPKIPLTTPSPSPTAAAAPTTTTLSPPPTQQQPPQSAAPAPSPLLPQQQPTPSA
APAPSPLLPQQQPPPSAARAPSPLPPQQQPLPSATPAPPPAQQLPPSATTLEPEKNHP
PAADRAGTEISPSPPFGQQPSFGDDASGGSGLVRYLSDLEEPFLSMSDSEEAESDLAS
DIPTTEDEDMFEDEVFSNSLESGSSAPTSPITLDTARSQYYQTTFDIETPEMDFVPLE
SNIARIAGHTYQEQAIVYDPASNREVPEADALSMIDYLLVTVVLEQGLIRSRDRSSVL
NLLEFLKDWSGHLQVPTLDLEQLLTSELNIQNLANMLSENKGRAGEFHKHLAAKLEAC
LPSLATKDAVRVDAGAKMLAEIPQLAESDDGKFDLEAARRRLTDLLSGGDQEAGEGGG
EPEDNSIYRGPHVDVPLVLDDESWKRLLSLAEAARTAVARQQAGVDEEDVRFLALLTA
IEYGAPPAASVPPFVHNVAVRSKNAALHVRRCTADIRDKVASAASDYLSYLEDPSLPT
VMDFDDLLTHLRHTCQIIASLPLLNIRYTSIEWDYRELLYLGTALSDMSGIPWPLERV
EEDDPSIAPLPEFETVAKKQKELETTRENEKRLRTILDDIEAMLGLAGVASAPGAPIS
PASPSATPANHDNPEATPPLADTAALTIPVIEKYIANAGSIVGAAKNPTYIRLRDTIQ
QIVRSKKYLMNILKSITFYTIDNYIASFEESIDHLYRDLPVLDPEVQDGIDRILDPMV
SEALHTFEMGNRLTLEPARLVALQNFATHSTLKETAAAVNLLPGLLAVYDATITGQAP
EDALRLLSGLQNQLSQTLIPGKLKKRFLSYLQKLKNNNNDQLRQKEVQAWRLEAEGFK
PATEEQLEAFLDTAPNKELKRQYEKKLRQLMETGRKEKEKLREQEDKERQERRAREAN
EAWARIRKALGARPEPAPTSPDDWNTLLASLLPDNTDSAAAAAAAVARNTDILDSLTQ
ILAAMLLGITRVRRERLRSLLVDDGGAAERMEAAEPGWFTDIETGPLARLDAWPATPA
ATAKEGGGGRGAEEAAGALFRARTAADAIRSALAQTRQALQSPDMKSAVVNTDLEAPY
AEYERGLAGLLEKRRAAEAALTAIVSEYVDRTLPEATNDPGQANLPPPPTIPQATAPP
RLASDSALWPKKPQLLTRRERDDLLQATGDFFSELLTEAEAAEVRALEEQVRESQTLM
AKAHEMAASTRRGFHTALEAVLSRSRDEAPDDELRSLLPSPPKAPVQAPLEAALARAA
AGNGSWPYRKSLAAAKWIRGICEAVRGLSEGALALAGGAGAWLNLAAAADGEIHELTR
LLEVEGMAQNSMDGMEELRLALATLDPKRVAGGKETVADWKRRLSRLEAIIQEAQEES
QLQGTLQDLVTQARGHTDPRQLKIVVEAARGLALGASAGSQYALLKDKLLRYASAKQS
FLAFYETAQPTVFVKHPLTNNLPLLITISAPPTGWGNGAPTRRAQFLAAAGPAKYAGT
LWLETESPCDPLNPAYVSADTQEPLNYIPVYHNFLEYVMPTVLENPEAFSLTPAGRPQ
AIGPPQDDQERRRRTLASVASARLSAAAADSYWDTWPDVESNAGELLREYVSAPKALM
EDLADNPIVAMTLLAHASLIASRNHPPYPAPATDREVILLEQREMMALLVGTHPAYAA
AFLGAPSFYAGLGLVSALARDGGLGDLLSDSVLTYRLVRSPASGRGGMPSTTRGSNDG
EDARRLTRHRIAGPPTGFIFFQDAWEEMDTRAALWPHPEFLGLVHNQSTARARACMLL
LARRCFAPEALQQLWHSLRPLEGPVAFQDYLRDFVKQAYTRGEELPRAEGLEVPRETP
SSYGTVTGRALRNLMPYGTPITGPKRGSGDTIPVSVFEAAVAAAFLGRPLTLFVSSQY
LFNLKTLGQVRVVAPLLYCDGHSEPFRSLVETISLNFLQDLDGYSESFEPEMSIFARQ
AVWLRELLTEARAAKPKEARPPTVAILANRKNIIWKCFTYRHNLPDVQFYFNAAGASR
WPTDVLNPSFYEHEDPPLPVGYQLPPNPRNVQELFSGFPPRVGHGLVSGDGFQSADNT
PASSDRLQQLGGGETDQGEKGSTTAESEASGPPSPQSPLLEKVAPGRPRDWLSPTSSP
RDVTVTPGLAAPITLPGPRLMARPYFGAETRASESPDRSPGSSPRPWPKDSLELLPQP
APQQPPSSPWASEQGPIVYTLSPHSTPSTASGSQKKHTIQIPGLVPSQKPSYPPSAPY
KPGQSTGGIAPTPSAASLTTFGLQPQDTQASSQDPPYGHSIMQREKKQQGGREEAAEI
RPSATRLPTAVGLRPRAPVVAAGAAASATPAFDPGEAPSGFPIPQAPALGSGLAAPAH
TPVGALAPRPQKTQAQRPQDAAALPTPTIKAVGARPVPKATGALAAGARPRGQPTAAP
PSAASPPRVSLPVRSRQQQSPAIPLPPMHSGSEPGARPEVRLSQYRHAGPQTYTVRKE
APPSAASQLPKMPKCKDSMYYPPSGSARYPAPFQALSFSQSVASPAPSSDQTTLLWNT
PSVVTQFLSIEDIIREVVTGGSTSGDLVVPSGSPSSLSTAAPEQDLRYSLTLSQASRV
LSRFVSQLRRKLERSTHRLIADLERLKFLYL"

misc_feature 62249
/note="BAM: BamH1 F/Q"

misc_feature 62430. .62477
/note="Site III for EBNA-1 binding (Rawlins et al, 1985)"

misc_feature 66121
/note="BAM: BamH1 Q/U"

mRNA 67477. .67649
/note="Exon in EBNA-1 RNA (Speck and Strominger,1985) and cDNA clone T4 (Bodescot et al, 1986)"

misc_feature 69410
/note="BAM: BamH1 U/P"

repeat_region 69684. .69930
/note="5 x 51bp repeats"

repeat_region 70387. .70521
/note="9 x 15bp repeat"

promoter 70750
/note="TATA: CATAAAA"

CDS complement(71520. .75239)
/note="BOLF1 reading frame, 1 NXT/S analogous to VZV RF 21"
/codon_start=1
/db_xref="PID:g1334855"
/db_xref="SWISS-PROT:P03189"
/translation="MASAMESDSSGGSGGADAQPPLAEVDGGLARVTRQLLLSGDDPA
ARLRALMPLELGIFGLGDLAQPVLVRDFLNTLTLMSGHAYPAAVLRHHAYYLLRAASF
SRRSFGLGHLEAALDVLASSLPPTTASPATDDPLDGSRLIAETRALAAAYRRIIEEGS
GEVLAVSGPTATFAFVEELVADTYLARWDAFPREGLSFYAFNAAKTTLGRWLVTVYAE
TNRYPWAAAGQGQPTAADIKAMAVELVEHSGGGAGGGEGEESGGGLFHRPESLSSVVA
SLPLARRRAVEILGVYAEASGGQTPPVAAVPVLAFDAARLRLLEPSGALFYDYVYEAL
LWDQTYGVPDSVIEAFLAGMAAEMEALAARVQEAAGSRASFSPAAIEQVATVLLSAGL
NETVAGDYAMMLASVPRVSRSRWRWLEATAALLESLSGFALHFFRLLPTASPTSRFAR
VARAAYLRAEAEAVDRRARRTSGPSTPAAAPAATAVGVGAAADPWDAVTPLRIFIVPP
PAAEYEQVAGDLSSELLRSLLWVRYSRLWQAPAPAPALPCKPPLLPGEQGRRQWTAAV
AAAPRTDVEAYCRSLRAGQTARADPAYVHSPFFPAAFIEFQIWPALRRVLSNELPKTR
SLAALRWLVSFGSDLALPSPELTRARRPLELIYATVWEIYDGAPPMPGESPQAVGLRP
LNLEGEGKAGDAGAEGAEDEEGGGPWGLSSHDAVLRIMDAVREVSGIISETISASERA
AEAPPLAWPTSLFSLLFTLRYSTTAESLGLATRRFLVSGETLSEDISRLTGAAWRLCS
RPLLYDAETGRVQIPLATEEEEEEAVVAVKEKSVSSSPRHYSTDLQTLKSVVEGIQDVC
RDAAARWALATADTATLRRRLLVPALRESRGIADHPLWAHTSEPLRPDLEELNERVEH
ALELGYSLTGALRRSVAYRFRDYTFARLFQPPAIDAERAEAIVRRDARPPPVFIPAPR
RLPQGGADTPPPLSMDDILYLGKSICKALVDVLDHHPAAPETTPIKTYTPAMDLNPEQ
ITVTPRSPSVLAAFARTARVQTHHLVPALTDDSPSPVGQTPPPFRILPAKKLAAILLG
NGRNASKRRASRDLSPPPHGRWRAVLDSSPFSFSSSDFSDQDEGEGGEADLRGVPGGG
GEGAYEEDRERPSDIDTAARAQKVETSCPRRRSPRTTPSPSRRASGGGGPDRGEAEAH
TYPPYLSAAAAASRVRPRTRRGATRRPPRPTAEDE"

promoter complement(72192)
/note="TATA: TATTAAA before BPLF1"

misc_feature 73468
/note="BAM: BamH1 P/O"

promoter 75017
/note="TATA: TATTTAA BO-R1 late promoter before BORF1, gives 3.9kb late RNA"

CDS 75238. .76332
/note="BORF1 late reading frame, 2 NXT/S homologous to VZV RF20"
/codon_start=1
/db_xref="PID:g1334854"
/db_xref="SWISS-PROT:P03187"

```
                /translation="MKVQGSVDRRRLQRRIAGLLPPPARRLNISRGSEFTRDVRGLVE
                EHAQASSLSAAAVWRAGLLAPGEVAVAGGGSGGGSFSWSGWRPPVFGDFLIHASSFNN
                AEATGTPLFQFKQSDPFSGVDAVFTPLSLFILMNHGRGVAARVEAGGGLTRMANLLYD
                SPATLADLVPDFGRLVADRRFHNFITPVGPLVENIKSTYLNKITTVVHGPVVSKAIPR
                STVKVTVPQEAFVDLDAWLSGGAGGGGGVCFVGGLGLQPCPADARLYVALTYEEAGPR
                FTFFQSSRGHCQIMNILRIYYSPSIMHRYAVVQPLHIEELTFGAVACLGTFSATDGWR
                RSAFNYRGSSLPVVEIDSFYSNVSDWEVIL"
promoter        complement(75322)
                /note="TATA: TATTTAG before BOLF1"
promoter        75819
                /note="TATA: TATAAAG"
misc_feature    75838
                /note="polyA signal: AATAAA"
misc_feature    complement(76126)
                /note="polyA signal: AATAAA"
promoter        76169
                /note="TATA: TACATAT BO-R2 early promoter before BORF2,
                gives 2.8kb RNA"
misc_feature    complement(76300)
                /note="polyA signal: AATAAA"
CDS             76407. .78887
                /note="BORF2 early reading frame, 2 NXT/S. Homology HSV
                140K ribonucleotide reductase (Gibson et al, 1984) and RF
                19 VZV"
                /codon_start=1
                /db_xref="PID:g1334856"
                /db_xref="SWISS-PROT:P03190"
                /translation="MATTSHVEHELLSKLIDELKVKANSDPEADVLAGRLLHRLKAES
                VTHTVAEYLEVFSDKFYDEEFFQMHRDELETRVSAFAQSPAYERIVSSGYLSALRYYD
                TYLYVGRSGKQESVQHFYMRLAGFCASTTCLYAGLRAALQRARPEIESDMEVFDYYFE
                HLTSQTVCCSTPFMRFAGVENSTLASCILTTPDLSSEWDVTQALYRHLGRYLFQRAGV
                GVGVTGAGQDGKHISLLMRMINSHVEYHNYGCKRPVSVAAYMEPWHSQIFKFLETKLP
                ENHERCPGIFTGLFVPELFFKLFRDTPWSDWYLFDPKDAGDLERLYGEEFEREYYRLV
                TAGKFCGRVSIKSLMFSIVNCAVKAGSPFILLKEACNAHFWRDLQGEAMNAANLCAEV
                LQPSRKSVATCNLANICLPRCLVNAPLAVRAQRADTQGDELLLALPRLSVTLPGEGAV
                GDGFSLARLRDATQCATFVVACSILQGSPTYDSRDMASMGLGVQGLADVFADLGWQYT
                DPPSRSLNKEIFEHMYFTALCTSSLIGLHTRKIFPGFKQSKYAGGWFHWHDWAGTDLS
                IPREIWSRLSERIVRDGLFNSQFIALMPTSGCAQVTGCSDAFYPFYANASTKVTNKEE
                ALRPNRSFWRHVRLDDREALNLVGGRVSCLPEALRQRYLRFQTAFDYNQEDLIQMSRD
                RAPFVDQSQSHSLFLREEDAARASTLANLLVRSYELGLKTIMYYCRIEKAADLGVMEC
                KASAALSVPREEQNERSPAEQMPPRPMEPAQVAGPVDIMSKGPGEGPGGWCVPGGLEV
                CYKYRQLFSEDDLLETDGFTERACESCQ"
misc_feature    77835
                /note="BAM: Bam H1 O/a"
promoter        78804
                /note="TATA: TATAAGT Ba-R1 early promoter before BaRF1,
                gives 3.5kb RNA"
misc_feature    78883
                /note="polyA signal: AATAAA, end of 3.9kb late RNA from
                75017 and 2.8kb early RNA from 76169"
misc_feature    complement(78896)
                /note="polyA signal: AATAAA"
CDS             78900. .79808
                /note="BaRF1 early reading frame, homologous to HSV 38K
                ribonucleotide reductase (Gibson et al, 1984) and RF 18
                VZV"
                /codon_start=1
                /db_xref="PID:g1334857"
                /db_xref="SWISS-PROT:P03175"
                /translation="MSKLLYVRDHEGFACLTVETHRNRWFAAHIVLTKDCGCLKLLNE
                RDLEFYKFLFTFLAMAEKLVNFNIDELVTSFESHDIDHYYTEQKAMENVHGETYANIL
                NMLFDGDRAAMNAYAEAIMADEALQAKISWLRDKVAAAVTLPEKILVFLLIEGIFFIS
                SFYSIALLRVRGLMPGICLANNYISRDELLHTRAASLLYNSMTAKADRPRATWIQELF
                RTAVEVETAFIEARGEGVTLVDVRAIKQFLEATADRILGDIGQAPLYGTPPPKDCPLT
                YMTSIKQTNFFEQESSDYTMLVVDDL"
promoter        complement(79495)
                /note="TATA: TATAACA"
misc_feature    79537
                /note="BAM: Bam H1 a/M"
promoter        79840
                /note="TATA: CATAAAT BM-R1 early promoter before BMRF1,
                gives 2.5kb RNA"
CDS             79899. .81113
                /note="BMRF1 early reading frame. Early antigen protein
                recognised by R3 monoclonal (Pearson et al 1983; Cho et
                al, 1985a)"
                /codon_start=1
                /db_xref="PID:g1334858"
                /db_xref="SWISS-PROT:P03191"
                /translation="METTQTLRFKTKALAVLSKCYDHAQTHLKGGVLQVNLLSVNYGG
                PRLAAVANAGTAGLISFEVSPDAVAEWQNHQSPEEAPAAVSFRNLAYGRTCVLGKELF
                GSAVEQASLQFYKRPQGGSRPEFVKLTMEYDDKVSKSHHTCALMPYMPPASDRLRNEQ
                MIGQVLLMPKTASSLQKWARQQGSGGVKVTLNPDLYVTTYTSGEACLTLDYKPLSVGP
                YEAFTGPVAKAQDVGAVEAHVVCSVAADSLAAALSLCRIPAVSVPILRFYRSGIIAVV
                AGLLTSAGDLPLDLSVILFNHASEEAAASTASEPEDKSPRVQPLGTGLQQRPRHTVSP
                SPSPPPPPRTPTPWESPARPETPSPAIPSHSSNTALERPLAVQLARKRTSSEARQKQKH
                PKKVKQAFNPLI"
promoter        80779
                /note="TATA: TATTTAA BM-R2 late promoter before BMRF2"
misc_feature    complement(80782)
                /note="polyA signal: AATAAA"
promoter        80832
                /note="TATA: GATAAAA, possible promoter for 1.4kb late RNA
                encoding BMRF2"
CDS             81118. .82191
                /note="BMRF2 early reading frame"
                /codon_start=1
                /db_xref="PID:g1334859"
                /db_xref="SWISS-PROT:P03192"
                /translation="MFSCKQHLSLGACVFCLGLLASTPFIWCFVFANLLSLEIFSPWQ
                THVYRLGFPTACLMAVLWTLVPAKHAVRAVTPAIMLNIASALIFFSLRVYSTSTWVSA
                PCLFLANLPLLCLWPRLAIEIVYICPAIHQRFFELGLLLACTIFALSVVSRALEVSAV
                FMSPFFIFLALGSGSLAGARRNQIYTSGLERRRSIFCARGDHSVASLKETLHKCPWDL
                LAISALTVLVVCVMIVLHVHAEVFFGLSRYLPLFLCGAMASGGLYLGHSSIIACVMAT
                LCTLTSVVVYFLHETLGPLGKTVLFISIFVYYFSGVAALSAAMRYKLKKFVNGPLVHL
                RVVYMCCFVFTFCEYLLVTFIKS"
promoter        81751
                /note="TATA: CATAAAT"
misc_feature    82180
                /note="polyA signal: ATTAAA, end of 3.5kb early RNA from
                78804, 2.5kb early RNA from 79840 and 1.4kb late RNA"
promoter        complement(82311)
                /note="TATA: CATAAAT"
repeat_region   82319. .82461
                /note="2x71bp repeats"
CDS             complement(82743. .84059)
                /note="BMLF1 early reading frame. Diffuse early antigen
                (Cho et al, 1985b). Also homologous to RF 4 VZV and IE63
                of HSV.(BSLF2 + BMLF1) is also called EB2
                (Chevallier-Greco et al, 1986). General transactivator of
                transcription (Lieberman et al, 1986)."
                /codon_start=1
                /db_xref="PID:e25042"
                /db_xref="PID:g1632790"
                /db_xref="SWISS-PROT:Q04360"
                /translation="MEGSEEHSTDGEISSSEEEDEDPTPAHAIPARPSSVVITPTSAS
                FVIPRKKWDLQDKTVTLHRSPLCRDEDEKEETGNSSYTRGHKRRRGEVHGCTDESYGK
                RRHLPPGARAPRAPRAPRVPRAPRSPRAPRSNRATRGPRSESRGAGRSTRKQARQERS
                QRPLPNKPWFDMSLVKPVSKITFVTLPSPLASLTLEPIQDPFLQSMLAVAAHPEIGAW
                QKVQPRHELRRSYKTLREFFTKSTNKDTWLDARMQAIQNAGLCTLVAMLEETIFWLQE
                ITYHGDLPLAPAEDILLACAMSLSKVILTKLKELAPCFLPNTRDYNFVKQLFYITCAT
                ARQNKVVETLSSSYVKQPLCLLAAYAAVAPAYINANCRRRHDEVEFLGHYIKNYNPGT
                LSSLLTEAVETHTRDCRSASCSRLVRAILSPGTGSLGLFFVPGLNQ"
misc_feature    complement(82747)
                /note="polyA signal: AATAAA"
repeat_region   83640. .83729
                /note="10x9bp repeats"
misc_feature    complement(84122)
                /note="ACCEPT: CTCCCCTCTGCAG acceptor in spliced form of
                BMLF1 RNA"
misc_feature    complement(84227)
                /note="DONOR: CAGGTAAGA donor in spliced form of BMLF1
                RNA"
CDS             complement(84229. .84288)
                /note="BSLF2 early reading frame in 5' exon of spliced RNA
                encoding BMLF1"
```

```
                /codon_start=1
                /db_xref="PID:g1334861"
                /db_xref="SPTREMBL:Q66539"
                /translation="MVPSQRLSRTSSISSNEDPA"
misc_feature    84233
                /note="BAM: Bam H1 M/S"
CDS             complement(84257. .86881)
                /note="BSLF1 reading frame, homologous to RF 6 VZV"
                /codon_start=1
                /db_xref="PID:g1334862"
                /db_xref="SWISS-PROT:P03193"
                /translation="MSAPVVIKALVASNTDIAEAILDAILSRPDEGFRLFCLCHNASP
                LHHVAGSLVELQLHLPKKRLTSQSRCGLVLTLHLPAEEAFPFLRGLTPLTADRLSTYL
                DRAGALRSLTPLVELLTLSAKKQPQGDARGRVAWLRPKIVGCLRRIYRVNISARWFIS
                TFGSHEAQFVLVTAAYYFWGIPCTIETLAHLTELFTSESGQSLAAVTSLAELGEVFGS
                SAWAEQTEAFAHFAHEKLRRDSREIRAVARTIDAYRGRLPLASADLVRYVYLAHAQCF
                NEGTFKRYSQLTSMGEIGCLPSGGVVLPSLLDRGFAEHMRTYFTRETYLAEHVRVQQL
                KIRMEPPAPYTWDPDPDDGLMRAWAGLSVDVARELVELARWHADEGPTYPPTLQGFLC
                LAGQATCRGQWNPKEQFLPPTVLRRVQRLPVFLCHFADRHYFVMTAADPFSSHLAEVV
                STPTNCRLPDTCLTRALSYTPVYYSQNSLSEQLFVSRHEYFNPRLPVCNLVLDLDLKI
                KGAPWSLEEIYDLCRTVRREVLRLMRRLGPVSRAHPVYFFKSACPPADPDNMEDVLPF
                CICTGKLGFRVITPLPRGHAIVGTSAVQGFVSVLQKLMGLTACLRRMRHKIKEIGAPL
                FDSGVYHAGRCIRLPHTYKVDRGGGLSRQLRLFVCHPEEEDKHSYVKNALNIQNLLHH
                SLHVGWPAPKTFCYHIADDGRDYLIQRTRETLPPTVENVCAMIEGHLGLDLVAWVSSC
                IWPSLMSTLATAVPEDKFPQFLHVTFEQTGPNLVQVCHARGRNFACLRHTHRASSKNV
                RVFLVLYYTSQAITVTFMSQCFAGRCGANQPTAHFSISVPASRIINRAEASQDSTTSQ
                LARRRDRQDGSFSETLPN"
promoter        complement(84356)
                /note="TATA: CATAAAT before BSLF2 and BMLF1. Two RNAs
                start here; one is spliced and the other is unspliced,
                both traverse BMLF1."
promoter        86882
                /note="TATA: TATTTAA BS-R1 late promoter before BSRF1"
CDS             86924. .87580
                /note="BSRF1 reading frame"
                /codon_start=1
                /db_xref="PID:g1334863"
                /db_xref="SWISS-PROT:P03194"
                /translation="MAFYLPDWSCCGLWLFGRPRNRYSQLPEEPETFECPDRWRAEID
                LGLPPGVQVGDLLRNEQTMGSLRQVYLLAVQANSITDHLKRFDAVRVPESCRGVVEAQ
                VAKLEAVRSVIWNTMISLAVSGIEMDENGLKALLDKQAGDSLALMEMEKVATALKMDE
                TGAWAQEISAVVSSVTAPSASAPFINSAFEPEVPTPVLAPPPVVRQPEHSGPTELALT
                "
misc_feature    complement(87134)
                /note="polyA signal: AATAAA"
misc_feature    87599
                /note="polyA signal: AATAAA"
misc_feature    complement(87613)
                /note="polyA signal: AATAAA, end 1.0kb early RNA from
                BLL3"
CDS             complement(87638. .88474)
                /note="BLLF3 early reading frame (BLLF2 in Baer et al,
                1984). Homologous to RF 8 VZV and dUTPase HSV."
                /codon_start=1
                /db_xref="PID:g1334864"
                /db_xref="SWISS-PROT:P03195"
                /translation="MEACPHIRYAFQNDKLLLQQASVGRLTLVNKTTILLRPMKTTTV
                DLGLYARPPEGHGLMLWGSTSRPVTSHVGIIDPGYTGELRLILQNQRRYNSTLRPSEL
                KIHLAAFRYATPQMEEDKGPINHPQYPGDVGLDVSLPKDLALFPHQTVSVTLTVPPPS
                IPHHRPTIFGRSGLAMQGILVKPCRWRRGGVDVSLTNFSDQTVFLNKYRRFCQLVYLH
                KHHLTSFYSPHSDAGVLGPRSLFRWASCTFEEVPSLAMGDSGLSEALEGRQGRGFGSS
                GQ"
misc_feature    87650
                /note="BAM: Bam H1 S/L"
promoter        88507
                /note="TATA: TATATAT BL-R1 late promoter before BLRF1,
                gives 1.0kb late RNA"
promoter        88511
                /note="TATA: TATAAGA"
promoter        complement(88514)
                /note="TATA: TATATAT BL-L3 early promoter before BLLF3,
                gives 1.0kb early RNA"
CDS             88547. .88855
                /note="BLRF1 late reading frame"
                /codon_start=1
                /db_xref="PID:g1334865"
                /db_xref="SWISS-PROT:P03196"
                /translation="MGKVLRKPFAKAVPLLFLAATWLLTGVLPAGASSPTNAAAASLT
                EAQDQFYSYTCNADTFSPSLTSFASIWALLTLVLVIIASAIYLMYVCFNKFVNTLLTD
                "
promoter        88863
                /note="TATA: TATTTAA BL-R2 late promoter before BLRF2,
                gives 0.6kb late RNA"
CDS             88925. .89413
                /note="BLRF2 late reading frame, 2 NXS/T"
                /codon_start=1
                /db_xref="PID:g1334866"
                /db_xref="SWISS-PROT:P03197"
                /translation="MSAPRKVRLPSVKAVDMSMEDMAARLARLESENKALKQQVLRGG
                ACASSTSVPSAPVPPPEPLTARQREVMITQATGRLASQAMKKIEDKVRKSVDGVTTRN
                EMENILQNLTLRIQVSMLGAKGQPSPGEGTRPRESNDPNATRRARSRSRGREAKKVQI
                SD"
misc_feature    89412
                /note="polyA signal: AATAAA, end of 1.0kb and 0.6kb late
                RNAs"
misc_feature    complement(89425)
                /note="polyA signal: AATAAA, end of 0.7kb early, 2.2kb
                late and 2.8kb late RNA"
CDS             complement(89430. .92153)
                /note="BLLF1a, late reading frame, gp350 membrane antigen,
                36 NXT /S (Hummel et al, 1984; Biggin et al, 1984; Beisel
                et al, 1985)"
                /codon_start=1
                /db_xref="PID:g1334869"
                /db_xref="SWISS-PROT:P03200"
                /translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
                DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
                ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
                SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
                STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTS
                STSPVTTPTPNATSPTPAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPT
                PNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQAN
                ATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSD
                NSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGPGNSSTST
                KPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTE
                PTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRF
                SNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"
CDS             complement(89430. .92153)
                /note="BLLF1b, late reading frame gp220 membrane antigen,
                spliced form of BLLF1a (Hummel et al, 1984; Biggin et al,
                1984; Beisel et al, 1985)"
                /codon_start=1
                /db_xref="PID:g1334868"
                /db_xref="SWISS-PROT:P03200"
                /translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
                DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
                ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
                KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
                SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
                NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
                MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
                VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
                STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTS
                STSPVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPT
                PNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQAN
                ATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSD
                NSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGPGNSSTST
                KPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTE
                PTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRF
                SNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"
promoter        complement(89434)
                /note="TATA: TATAAAG"
```

```
CDS             complement(89567. .90013)
                /note="BLLF2 early reading frame (BLLF3 in Baer et al,
                1984)"
                /codon_start=1
                /db_xref="PID:g1334867"
                /db_xref="SWISS-PROT:P03199"
                /translation="MCPPVRQHPAQAPPAKRQALETVPHPQNRGRLMSPKARPPKMQR
                RPRPPVAKRRRFPRSPQQVERPILPPVESTPQDMEPGQVQSPPQITAVIQLRQDRDTM
                RPPIYLPALLANCGPAGLLRAHRLPQPKPPCQSRQRPSPDSQTSPC"
promoter        complement(90051)
                /note="TATA: TATAACA BL-L2 early promoter before BLLF2,
                gives 0.7kb early RNA"
intron          complement(90062. .90652)
                /note="intervening sequence in gp220 gene"
repeat_region   90177. .90639
                /note="21 copies of 21bp approximate repeat"
promoter        complement(92192)
                /note="TATA: TATTAAA BL-L1 late promoter before BLLF1a,b.
                Gives 2.8 and 2.2kb late RNAs"
mRNA            92238. .92581
                /note="Exon in Bodescot et al (1986) RNA (spliced from
                20763 to 92670)"
CDS             92243. .92602
                /note="BLRF3 reading frame"
                /codon_start=1
                /db_xref="PID:g1632791"
                /db_xref="SWISS-PROT:P03202"
                /translation="MDKDRPGPPALDDNMEEEVPSTSVVQEQVSAGDWENVLIELSDS
                SSEKEAEDAHLEPAQKGTKRKRVDHDAGGSAPARPMLPPQPDLPGREAILRRFPLDLR
                TLLQAIGAAATVSIPMA"
mRNA            92670. .95248
                /note="Exon in (Bodescot et al, 1986) RNA from 92581, to
                3' end"
CDS             92670. .95162
                /note="BERF1 frame, homology with BERF2b and BERF4. A
                fusion of BLRF3 with BERF1 encodes EBNA-3A, latent cycle
                gene. (Hennessy et al, 1986, Joab et al, 1987); Protein
                sequence is in conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25053"
                /db_xref="PID:g1334871"
                /db_xref="SWISS-PROT:P12977"
                /translation="RIDTRAIDQFFGSQISNTEMYIMYAMAIRQAIRDRRRNPASRRD
                QAKWRLQTLAAGWPMGYQAYSSWMYSYTDHQTTPTFVHLQATLGCTGGRRCHVTFSAG
                TFKLPRCTPGDRQWLYVQSSVGNIVQSCNPRYSIFFDYMAIHRSLTKIWEEVLTPDQR
                VSFMEFLGFLQRTDLSYIKSFVSDALGTTSIQTPWIDDNPSTETAQAWNAGFLRGRAY
                GIDLLRTEGEHVEGATGETREESEDTESDGDDEDLPCIVSRGGPKVKRPPIFIRRLHR
                LLLMRAGKRTEQGKEVLEKARGSTYGTPRPPVPKPRPEVPQSDETATSHGSAQVPEPP
                TIHLAAQGMAYPLHEQHGMAPCPVAQAPPTPLPPVSPGDQLPGVFSDGRVACAPVPAP
                AGPIVRPWEPSLTQAAGQAFAPVRPQHMPVEPVPVPTVALERPVYPKPVRPAPPKIAM
                QGPGETSGIRRARERWRPAPWTPNPPRSPSQMSVRDRLARLRAEAQVKQASVEVQPPQ
                LTQVSPQQPMEGPLVPEQQMFPGAPFSQVADVVRAPGVPAMQPQYFDLPLIQPISQGA
                PVAPLRASMGPVPPVPATQPQYFDIPLTEPINQGASAAHFLPQQPMEGPLVPEQWMFP
                GAALSQSVRPGVAQSQYFDLPLTQPINHGAPAAHFLHQPPMEGPWVPEQWMFQGAPPS
                QGTDVVQHQLDALGYTLHGLNHPGVPVSPAVNQYHLSQAAFGLPIDEDESGEGSDTSE
                PCEALDLSIHGRPCPQAPEWPVQEEGGQDATEVLDLSIHGRPRPRTPEWPVQGEGGQN
                VTGPETRRVVVSAVVHMCQDDEFPDLQDPPDEA"
misc_feature    92703
                /note="BAM: Bam H1 L/E"
promoter        complement(93161)
                /note="TATA: CATAAAT"
promoter        93479
                /note="TATA: TATAAGA"
promoter        complement(93482)
                /note="TATA: TATAAAT"
repeat_region   94208. .94277
                /note="repeat type A"
repeat_region   94281. .94306
                /note="repeat type B"
repeat_region   94307. .94381
                /note="repeat type C"
repeat_region   94386. .94411
                /note="repeat type B"
repeat_region   94412. .94489
                /note="repeat type C"
repeat_region   94490. .94560
                /note="repeat type A"
repeat_region   94571. .94648
                /note="repeat type C"
repeat_region   94649. .94719
                /note="repeat type A"
repeat_region   94896. .94982
                /note="repeat type D"
repeat_region   94983. .95069
                /note="repeat type D"
misc_feature    95221
                /note="polyA signal: AATAAA"
misc_feature    complement(95272)
                /note="polyA signal: AATAAA"
CDS             join(95353. .95709,95788. .98247)
                /codon_start=1
                /product="EBNA3B (EBNA4A) latent protein"
                /db_xref="PID:e276431"
                /db_xref="PID:g1632792"
                /db_xref="SWISS-PROT:P03203"
                /translation="MKKAWLSRAQQADAGGASGSEDPPDYGDQGNVTQVGSEPISPEI
                GPFELSAASEDDPQSGPVEENLDAAAREEEEPHEQEHNGGDDPLDVHTRQPRFVDVNP
                TQAPVIQLVHAVYDSMLQSDLRPLGSLFLEQNLNIEEFIWMCMTVRHRCQAIRKKPLP
                IVKQRRWKLLSSCRSWRMGYRTHNLKVNSFESGGDNVHPVLVTATLGCDEGTRHATTY
                SAGIVQIPRISDQNQKIETAFLMARRARSLSAERYTLFFDLVSSGNTLYAIWIGLGTK
                NRVSFIEFVGWLCKKDHTHIREWFRQCTGRPKAAKPWLRAHPVAIPYDDPLTNEEIDL
                AYARGQAMNIEAPRLPDDPIIVEDDDESEEIEAESDEEEDKSGMESLKNIPQTLPYNP
                TVYGRPAVFDRKSDAKSTKKCRAIVTDFSVIKAIEEEHRKKKAARTEQPRATPESQAP
                TVVLQRPPTQQEPGPVGPLSVQARLEPWQPLPGPQVTAVLLHEESMQGVQVHGSMLDL
                LEKDDEVMEQRVMATLLPPVPQQPRAGRRGPCVFTGDLGIESDEPASTEPVHDQLLPA
                PGPDPLEIQPLTSPTTSQLSSSAPSCAQTPWPVVQPSQTPDDPTKQSRPPETAAPRQW
                PMPLRPIPMRPLRMQPIPFNHPVGPTPHQTPQVEITPYKPTWAQIGHIPYQPTPTGPA
                TMLLRQWAPATMQTPPRAPTPMSPPEVPPVPRQRPRGAPTPTPPPQVPPVPRQRPRGA
                PTPTPPPQVLPTPMQLALRAPAGQQGPTKQILRQLLTGGVKKGRPSLKLQAALERQAA
                AGWQPSPGSGTSDKIVQAPIFYPPVLQPIQVMGQGGSPTAMAASAVTQAPTEYTRERR
                GVGPMPPTDIPPSKRAKIEAYTEPEMPHGGASHSPVVILENVGQGQQQTLECGGTAKQ
                ERDMLGLGDIAVSSPSSSETSNDE"
mat_peptide     95353. .95724
                /note="BERF2a reading frame"
mat_peptide     95725. .98244
                /note="BERF2b frame, homology with BERF1 and BERF4. BERF2a
                and BERF2b are spliced together to make EBNA3B (EBNA4A)
                latent protein."
misc_feature    complement(95819)
                /note="polyA signal: AATAAA"
promoter        complement(95853)
                /note="TATA: TATAAAT"
misc_feature    complement(96276)
                /note="polyA signal: AATAAA"
repeat_region   97522. .97698
                /note="3x60bp repeat"
mat_peptide     98323. .98769
                /note="BERF3 reading frame"
mRNA            98364. .98730
                /note="Exon in EBNA-1 RNA (Speck and Strominger, 1985)"
CDS             join(98371. .98730,98805. .101423)
                /codon_start=1
                /product="EBNA3C (EBNA 4B) latent protein"
                /db_xref="PID:e276455"
                /db_xref="PID:g1632793"
                /db_xref="SWISS-PROT:P03204"
                /translation="MESFEGQGDSRQSPDNERGDNVQTTGEHDQDPGPGPPSSGASER
                LVPEESYSRDQQPWGQSRGDENRGWMQRIRRRRRRRAALSGHLLDTEDNVPPWLPPHD
                ITPYTARNIRDAACRAVKQSHLQALSNLILDSGLDTQHILCFVMAARQRLQDIRRGPL
                VAEGGVGWRHWLLTSPSQSWPMGYRTATLRTLTPVPNRVGADSIMLTATFGCQNAART
                LNTFSATVWTPPHAGPREQERYAREAEVRFLRGKWQRRYRRIYDLIELCGSLHHIWQN
                LLQTEENLLDFVRFMGVMSSCNNPAVNYWFHKTIGNFKPYYPWNAPPNENPYHARRGI
                KEHVIQNAFRKAQIQGLSMLATGGEPRGDATSETSSDEDTGRQGSDVELESSDDELPY
                IDPNMEPVQQRPVMFVSRVPAKKPRKLPWPTPKTHPVKRTNVKTSDRSDKAEAQSTPE
                RPGPSEQSSVTVEPAHPTPVEMPMVILHQPPPVPKPVPVKPTPPPSRRRRGACVVYDD
                DVIEVIDVETTEDSSSVSQPNKPHRKHQDGFQRSGRRQKRAAPPTVSPSDTGPPAVGP
```

PAAGPPAAGPPAAGPPAAGPPAAGPPAAGPRILAPLSAGPPAAGPHIVTPPSARPRIM
APPVVRMFMRERQLPQSTGRKPQCFWEMRAGREITQMQQEPSSHLQSATQPTTPRPSW
APSVCALSVMDAGKAQPIESSHLSSMSPTQPISHEEQPRYEDPDAPLDLSLHPDVAAQ
PAPQAPYQGYQEPPAPQAPYQGYQEPPPPQAPYQGYQEPPAHGLQSSSYPGYAGPWTP
RSQHPCYRHPWAPWSQDPVHGHTQGPWDPRAPHLPPQWDGSAGHGQDQVSQFPHLQSE
TGPPRLQLSLVPLVSSSAPSWSSPQPRAPIRPIPTRFPPPPMPLQDSMAVGCDSSGTA
CPSMPFASDYSQGAFTPLDINATTPKRPRVEESSHGPARCSQATAEAQEILSDNSEIS
VFPKDAKQTDYDASTESELD"

misc_feature 98731
/note="DONOR: AAGGTGAGT donor"

mat_peptide 98805. .101420
/note="BERF4 frame, homology with BERF1 and BERF2b. BERF3 and BERF4 are spliced together to make the EBNA3C (EBNA 4B) latent protein."

mRNA 98805. .99050
/note="Exon in T4 cDNA (Bodescot et al 1986). 99050 is not the end of the RNA."

misc_feature 99126. .102118
/note="DEL: Deletion in Raji"

promoter 99443
/note="TATA: CATAAAA"

misc_feature 100104
/note="DONOR: ACCGTGAGT possible donor before repeat."

repeat_region 100122. .100304
/note="10 x 15bp repeat"

misc_feature 100528
/note="DONOR: CTGGTAAGG possible donor"

misc_feature 100613
/note="BAM: Bam H1 E/e1"

repeat_region 100665. .100781
/note="3x39bp repeat"

promoter complement(100860)
/note="TATA: TATAACA"

misc_feature 100919
/note="BAM: Bam H1 e1/e2"

misc_feature 101426
/note="BAM: Bam H1 e2/e3"

CDS complement(101445. .102116)
/note="BZLF2 reading frame 3x NXT/S. 2.5kb late RNA traverses BZLF2, ends unknown."
/codon_start=1
/db_xref="PID:g1334876"
/db_xref="SWISS-PROT:P03205"
/translation="MVSFKQVRVPLFTAIALVIVLLLAYFLPPRVRGGGRVAAAAITW
VPKPNVEVWPVDPPPPVNFNKTAEQEYGDKEVKLPHWTPTLHTFQVPQNYTKANCTYC
NTREYTFSYKGCCFYFTKKKHTWNGCFQACAELYPCTYFYGPTPDILPVVTRNLNAIE
SLWVGVYRVGEGNWTSLDGGTFKVYQIFGSHCTYVSKFSTVPVSHHECSFLKPCLCVS
QRSNS"

promoter 101690
/note="TATA: CATAAAA"

misc_feature 101765
/note="polyA signal: AATAAA"

promoter complement(101786)
/note="TATA: TATAAAG"

misc_feature 101947
/note="BAM: Bam H1 e3/Z"

misc_feature complement(102098)
/note="DONOR: CAGGTGAGG possible donor"

mRNA complement(102126. .102341)
/note="3' terminal exon of 0.9kb and 2.8kb early RNAs"

promoter 102153
/note="TATA: TATTAAT"

misc_feature complement(102156)
/note="polyA signal: AATAAA 3' end of 0.9kb and 2.8kb RNAs encoding BZLF1 and BRLF1"

promoter complement(102160)
/note="TATA: TATTAAT"

CDS complement(join(<102210. .102338,102423. .102530, 102655. .103155))
/note="BZLF1 reading frame, modified from Baer et al, 1984. Has two splices within frame. 2xNXT/S. Immediate early gene which disrupts latency (Countryman and Miller, 1985), called EB1 by Chevallier-Greco et al, 1986 and ZEBRA by Miller."
/codon_start=1
/db_xref="PID:e276432"
/db_xref="PID:g1632794"
/db_xref="SWISS-PROT:P03206"
/translation="MMDPNSTSEDVKFTPDPYQVPFVQAFDQATRVYQDLGGPSQAPL PCVLWPVLPEPLPQGQLTAYHVSTAPTGSWFSAPQPAPENAYQAYAAPQLFPVSDITQ NQQTNQAGGEAPQPGDNSTVQTAAAVVFACPGANQGQQLADIGVPQPAPVAAPARRTR KPQQPESLEECDSELEIKRYKNRVASRKCRAKFKQLLQHYREVAAAKSSENDRLRLLL KQMCPSLDVDSIIPRTPDVLHEDLLNF"

promoter complement(102380)
/note="TATA: CATAAAT"

promoter 102415
/note="TATA:TATATAC"

promoter complement(102420)
/note="TATA: TATATAC"

mRNA complement(102426. .102530)
/note="Exon of 0.9kb and 2.8kb early RNAs"

misc_feature complement(102504)
/note="polyA signal: AATAAA, apparently not functional"

repeat_region 102581. .102652
/note="semi-repetitive sequence, homologous to human c-fos 3' sequence"

mRNA complement(102655. .103194)
/note="First exon of 0.9kb early RNA encoding BZLF1"

misc_feature complement(102918)
/note="splice acceptor used in RZ fusion gene (Sargeant)"

promoter complement(103231)
/note="TATA: TTTAAA of BZL1 immediate early promoter gives 0.9kb RNA"

misc_feature complement(103256. .103311)
/note="Upstream of BZL1, homology to 106243 to 106188"

CDS complement(103366. .105183)
/note="BRLF1 reading frame, (immediate?) early gene, acts as transcription activator."
/codon_start=1
/db_xref="PID:g1334878"
/db_xref="SWISS-PROT:P03209"
/translation="MRPKKDGLEDFLRLTPEIKKQLGSLVSDYCNVLNKEFTAGSVEI TLRSYKICKAFINEAKAHGREWGGLMATLNICNFWAILRNNRVRRRAENAGNDACSIA CPIVMRYVLDHLIVVTDRFFIQAPSNRVMIPATIGTAMYKLLKHSRVRAYTYSKVLGV DRAAIMASGKQVVEHLNRMEKEGLLSSKFKAFCKWVFTYPVLEEMFQTMVSSKTGHLT DDVKDVRALIKTLPRASYSSHAGQRSYVSGVLPACLLSTKSKAVETPILVSGADRMDE ELMGNDGGASHTEARYSESGQFHAFTDELESLPSPTMPLKPGAQSADCGDSSSSSSDS GNSDTEQSEREEARAEAPRLRAPKSRRTSRPNRGQTPCPSNAAEPEQPWIAAVHQESD ERPIFPHPSKPTFLPPVKRKKGLRDSREGMFLPKPEAGSAISDVFEGREVCQPKRIRP FHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPVPQPLDPAPAVTPEASHLLEDP DEETSQAVKALREMADTVIPQKEEAAICGQMDLSHPPPRGHLDELTTTLESMTEDLNL DSPLTPELNEILDTFLNDECLLHAMHISTGLSIFDTSLF"

misc_feature complement(103453. .103462)
/note="TAATGAAATC sequence"

misc_feature 103741
/note="BAM: Bam H1 Z/g"

misc_feature 103816
/note="BAM: Bam H1 g/R"

mRNA complement(104926. .105185)
/note="exon in RZ fusion gene (Sargeant)"

mRNA complement(104927. .104989)
/note="BRLF2 poss. small 5' exon"

promoter 105016
/note="TATA: TATAAAT before BRRF1, possible promoter for 1.1 kb early RNA encoding BRRF1"

promoter complement(105019)
/note="TATA: TATAAAT before BRLF2"

CDS 105182. .106114
/note="BRRF1 early reading frame"
/codon_start=1
/db_xref="PID:g1334877"
/db_xref="SWISS-PROT:P03207"
/translation="MASSNRGNARPLKSFLHELYLKHYPEVGDVVHLLNTIGVDCDLP PSHPLLTAQRGLFLARVLQAVQQHKLLEDTIVPKILKKLAYFLELLSYYSPKDEQRDI AEVLDHLKTNRDLGLDDRLWALIRKLRQDRHHASVNVLMPGSDYTAVSLQYYDGISIG MRKVIADVCRSGYASMPSMTATHNLSHQLLMASGPSEEPCAWRGFFNQVLLWTVALCK

```
                FRRCIYYNYIQGSIATISQLLHLEIKALCSWIISQDGMRLFQHSRPLLTLWESVAANQ
                EVTDAITLPDCAEYIDLLKHTKHVLENCSAMQYK"
misc_feature    complement(105185)
                /note="ACCEPT: splice acceptor in 2.8kb early RNA encoding
                BRLF1 and RZ fusion gene (Sargeant)"
promoter        105213
                /note="TATA: CATTAAA"
misc_feature    106110
                /note="polyA signal: AATAAA, 3' end of early 1.1kb RNA
                encoding BRRF1"
misc_feature    complement(106125)
                /note="DONOR: CAGGTAAGA possible donor"
misc_feature    complement(106188. .106243)
                /note="Homology to upstream region of BZL1"
promoter        complement(106213)
                /note="TATA: CATAAAA"
promoter        106243
                /note="TATA: TATAAAA before BRRF2, possible promoter for
                1.8 kb RNA encoding BRRF2"
CDS             106302. .107915
                /note="BRRF2 reading frame"
                /codon_start=1
                /db_xref="PID:g1334879"
                /db_xref="SWISS-PROT:P03210"
                /translation="MSGQQRGSVILVPEHLAGALTKLMSDFITGQDVTLSGGNIAVKI
                RDAINQTPGGGDVAILSSLFALWNALPTSGRQSSRDDLIPAAVQALTTAHNLCLGVIP
                GETSHKDTPESLLRAIVTGLQKLWVDSCGCPECLQCLKGLKAIKPGLYEIPRIIPHTK
                QCSPVNLLNMLVHKLVALRGHVQLAYDARVLTPDFHEIPDLDDSDAVFARTLLAALFH
                LNMFFILKDYITQDSMSLKQALSGHWMSATGNPLPAAPETLRDYLEAFRNSDNHFYLP
                TTGPLNTFQFPEELLGRVVVIDSSLCAASHVQDVITHGVGAGVPRPRFSALPPAPSRE
                PQQTCSQLTSRGNESSRRNLGQPGGTSPAVPPVCPIVSLTASGAKQNRGGMGSLHLAK
                PEETSPAVSPVCPIASPAASRSKQHCGVTGSSQAAPSFSSVAPVASLSGDLEEEEEGS
                RESPSLPSSKKGDEEFEAWLEAQDANLEDVQREFSGLRVIGDEDEDGSEDGEFSDLDL
                SDSDHEGDEGGGAVGGGRSLHSLYSLSVV"
promoter        complement(106385)
                /note="TATA: GATAAAA"
misc_feature    complement(106973)
                /note="polyA signal: AATAAA"
promoter        complement(107124)
                /note="TATA: GATAAAA"
misc_feature    107457
                /note="BAM: Bam H1 R/f"
misc_feature    107565
                /note="BAM: Bam H1 f/K"
misc_feature    107914
                /note="polyA signal: AATAAA, 3' end of 1.8kb RNA encoding
                BBRF2"
misc_feature    107942
                /note="ACCEPT: splice acceptor for EBNA-1 RNA (from
                98730)"
CDS             107950. .109875
                /note="BKRF1 encodes EBNA-1 protein, latent cycle gene."
                /codon_start=1
                /db_xref="PID:g1334880"
                /db_xref="SWISS-PROT:P03211"
                /translation="MSDEGPGTGPGNGLGEKGDTSGPEGSGGSGPQRRGGDNHGRGRG
                RGRGRGGGRPGAPGGSGSGPRHRDGVRRPQKRPSCIGCKGTHGGTGAGAGAGGAGAGG
                AGAGGGAGAGGGAGGAGGAGGAGAGGGAGAGGGAGGAGGAGAGGGAGAGGGAGGAGAG
                GGAGGAGGAGAGGGAGAGGGAGGAGAGGGAGGAGAGGGAGAGGAGGAGAGGAGAGGA
                GAGGGAGGAGGAGAGGAGAGGAGAGGAGAGGAGGAGAGGAGGAGAGGAGGAGAGGGAG
                GAGAGGGAGGAGAGGAGGAGAGGAGGAGAGAGGAGGGAGAGGAGAGGGGRGRGGS
                GGRGRGGSGGRGRGGSGGRRGRGRERARGGSRERARGRGRGRGEKRPRSPSSQSSSSG
                SPPRRPPPGRRPFFHPVGEADYFEYHQEGGPDGEPDVPPGAIEQGPADDPGEGPSTGP
                RGQGDGGRRKKGGWFGKHRGQGGSNPKFENIAEGLRALLARSHVERTTDEGTWVAGVF
                VYGGSKTSLYNLRRGTALAIPQCRLTPLSRLPFGMAPGPGPQPGPLRESIVCYFMVFL
                QTHIFAEVLKDAIKDLVMTKPAPTCNIRVTVCSFDDGVDLPPWFPPMVEGAAAEGDDG
                DDGDEGGDGDEGEEGQE"
repeat_region   108217. .108924
                /note="EBNA triplet repeat GGA,GCA,GGG."
misc_feature    109856
                /note="DONOR: AGGGTGAGG possible donor at end BKRF1"
promoter        109905
                /note="TATA: TATTAAA before BKRF2, possible start for
                2.3kb late RNA"
misc_feature    109906
                /note="polyA signal: ATTAAA"
misc_feature    109937
                /note="polyA signal: AATAAA 3' end of EBNA-1 RNA"
CDS             109958. .110371
                /note="BKRF2 reading frame"
                /codon_start=1
                /db_xref="PID:g1334881"
                /db_xref="SWISS-PROT:P03212"
                /translation="MRAVGVFLAICLVTIFVLPTWGNWAYPCCHVTQLRAQHLLALEN
                ISDIYLVSNQTCDGFSLASLNSPKNGSNQLVISRCANGLNVVSFFISILKRSSSALTG
                HLRELLTTLETLYGSFSVEDLFGANLNRYAWHRGG"
misc_feature    110271
                /note="DONOR: TCCGTGAGT possible donor at end of BKRF2"
CDS             110353. .111120
                /note="BKRF3 reading frame, homologous to RF 59 VZV"
                /codon_start=1
                /db_xref="PID:e25020"
                /db_xref="PID:g1632795"
                /db_xref="SWISS-PROT:P12888"
                /translation="MASRGLDLWLDEHVWKRKQEIGVKGENLLLPDLWLDFLQLSPIF
                QRKLAAVIACVRRLRTQATVYPEEDMCMAWARFCDPSDIKVVILGQDPYHGGQANGLA
                FSVAYGFPVPPSLRNIYAELHRSLPEFSPPDHGCLDAWASQGVLLLNTILTVQKGKPG
                SHADIGWAWFTDHVISLLSERLKACVFMLWGAKAGDKASLINSKKHLVLTSQHPSPLA
                QNSTRKSAQQKFLGNNHFVLANNFLREKGLGEIDWRL"
misc_feature    111098
                /note="DONOR: TCGGTGAGA possible donor at end BKRF3"
CDS             111134. .111787
                /note="BKRF4 reading frame, contains complex repetitive
                sequence"
                /codon_start=1
                /db_xref="PID:e25021"
                /db_xref="PID:g1632796"
                /db_xref="SWISS-PROT:P30117"
                /translation="MAMFLKSRGVRSCRDRRLLSDEEEETSQSSSYTLGSQASQSIQE
                EDVSDTDESDYSDEDEEIDLEEEYPSDEDPSEGSDSDPSWHPSDSDESDYSESDEDEA
                TPGSQASRSSRVSPSTQQSSGLTPTPSFSRPRTRAPPRPPAPAPVRGRASAPPRPPAP
                VQQSTKDKGPHRPTRPVLRGPAPRRPPPPSSPNTYNKHMMETTPPIKGNNNYNWPWL"
misc_feature    111272
                /note="DONOR: GACGTGAGT poss.donor before rpt.seq. in
                BKRF4"
misc_feature    111719
                /note="polyA signal: AATAAA"
misc_feature    111787
                /note="polyA signal: AATAAA : currently unknown which is
                3' end of the 2.3kb late and 1.1kb early RNAs"
CDS             complement(111830. .114259)
                /note="BBLF4 early reading frame, very good homology to
                RF55 VZV"
                /codon_start=1
                /db_xref="PID:g1334885"
                /db_xref="SWISS-PROT:P03214"
                /translation="MAEEPRAPEALSSTFMLNMTSDASVRRIVRRIGTLARRRVQQLP
                DMETFSPEFDPELSEPPFLPFSAYVITGTAGAGKSTSVSCLHHTMDCLVTGATTVAAQ
                NLSQTLRAYCPTVYSAFGFKSRHINMTQRVSSHGRSTDAALEELQRRDLAKYWPVLSD
                IAAEFRRTKPRGLYSGVSGPAFEVLRDMHQGQLWTTNVIVVDEAGTLSVHILTAVVFC
                YWFFNAWLRTPLYRRGRIPCIVCVGSPTQTDAFQSSFSHETQVNKIRECDNILTFLVG
                NPRAATYVDVARNWALFINNKRCTDVQFGHLMKTLEYGLELSPDILAYVDRFVVPRAA
                IMDPAQYVGWTRLFLSHAEVKTFLTTLHATLKTAGQGRAARGTGGDGGGVTMFTCPVE
                CEVFLDPLAQYKTLVGLPGLTAHTWLQKNYARLGNYSQFADQDMVPVGTEQDEERVKV
                TYNVTYVKHSSVSVNCKTKKSICGYTGTFGDFMDTLEADSFVEAHGHEQPEYVYSFLA
                RLIYGGIYAFSHGGHSLCENGEYVAELGAVPLPGRTWDPEVTAGMELGELPLEVAWDG
                ERSPAAVFYARVLAPPAANSAPLCSLLNIYNDLRAYFRQCLDVAVRYGGREFRDLPFC
                TFTNNMLIRDNIEFTSDEPLLHGLLDYASTTENYTLLGYTHLNVFFGIRGKQQPQDAG
                SSRMPRLMVKDEAGFVCCLEHNTNKLYETIEDKSLNLCSIRDYGISSKLAMTIAKAQG
                LSLNKVAICFGSHRNIKPGHVYVALSRARHSNCVVMDRNPLSEMITGEGNPASGYIVD
                ALKNSRALLVY"
misc_feature    complement(111830)
                /note="polyA signal: AATAAA"
promoter        112471
                /note="TATA: TATATAT"
promoter        complement(112476)
```

```
                     /note="TATA: TATATAA"
misc_feature         112620
                     /note="BAM: Bam H1 K/B"
promoter             113876
                     /note="TATA: TATTTAT before BBRF1"
promoter             complement(113885)
                     /note="TATA: CATAAAT"
CDS                  114204. .116045
                     /note="BBRF1 late reading frame, homologous to RF 54 VZV"
                     /codon_start=1
                     /db_xref="PID:g1334884"
                     /db_xref="SWISS-PROT:P03213"
                     /translation="MFNMNVDESASGALGSSAIPVHPTPASVRLFEILQGKYAYVQGQ
                     TIYANLRNPGVFSRQVFTHLFKRAISHCTYDDVLHDWNKFEACIQKRWPSDDSCASRF
                     RESTFESWSTTMKLTVRDLLTTNIYRVLHSRSVLSYERYVDWICATGMVPAVKKPITQ
                     ELHSKIKSLRDRCVCRELGHERTIRSIGTELYEATKEIIESLNSTFIPQFTEVTIEYL
                     PRSDEYVAYYCGRRIRLHVLFPPAIFAGTVTFDSPVQRLYQNIFMCYRTLEHAKICQL
                     LNTAPLKAIVGHGGRDMYKDILAHLEQNSQRKDPKKELLNLLVKLSENKTISGVTDVV
                     EEFITDASNNLVDRNRLFGQPGETAAQGLKKKVSNTVVKCLTDQINEQFDQINGLEKE
                     RELYLKKIRSMESQLQASLGPGGNNPAASAPAAVAAEAASVDILTGSTASAIEKLFNS
                     PSASLGARVSGHNESILNSFVSQYIPPSREMTKDLTELWESELFNTFKLTPVVDNQGQ
                     RLYVRYSSDTISILLGPFTYLVAELSPVELVTDVYATLGIVEIIDELYRSSRLAIYIE
                     DLGRKYCPASATGGDHGIRQAPSARGDTEPDHAKSKPARDPPPGAGS"
CDS                  115948. .116784
                     /note="BBRF2 late reading frame, homologous to RF 53 VZV"
                     /codon_start=1
                     /db_xref="PID:e25024"
                     /db_xref="PID:g1632797"
                     /db_xref="SWISS-PROT:P29882"
                     /translation="MASGKHHQPGGTRSLTMQKVSLRVTPRLVLEVNRHNAICVATNV
                     PEFYNARGDLNIRDLRAHVKARMISSQFCGYVLVSLLDSEDQVDHLNIFPHVFSERMI
                     LYKPNNVNLMEMCALLSMIENAKSPSIGLCREVLGRLTLLHSKCNNLDSLFLYNGART
                     LLSTLVKYHDLEEGAATPGPWNEGLSLFKLHKELKRAPSEARDLMQSLFLTSGKMGCL
                     ARSPKDYCADLNKEEDANSGFTFNLFYQDSLLTKHFQCQTVLQTLRRKCLGSDTVSKI
                     IP"
promoter             complement(116683)
                     /note="TATA: GATAAAA"
misc_feature         complement(116696)
                     /note="polyA signal: AATAAA"
CDS                  complement(116784. .117386)
                     /note="BBLF3 early reading frame, spliced to BBLF2. BBLF3
                     contains a consensus nucleotide binding site; Protein
                     sequence is in conflict with the conceptual translation"
                     /codon_start=1
                     /db_xref="PID:e25025"
                     /db_xref="PID:g1334887"
                     /db_xref="SPTREMBL:Q04343"
                     /translation="AFLQGVKDSEDASRLDRDVMGGEATVARRHIRVKARRGPGCLLM
                     AIFQGDLYVGGCREHSGPFLVWHEAFSWTLDQLAARPEADKAPPSHDHLLTLVRDLTR
                     RLAPGRRRNRFWALPRAWLQRLRRAGLRLSGSHVCLLDKDGARPAPCQTATEHGLSPT
                     AYFREIMAFLLDVISALHPGYTIPMEITRETDLLMTVLSLF"
misc_feature         116785
                     /note="polyA signal: AATAAA"
intron               complement(117386. .117515)
                     /note="intron spliced out in RNA linking BBLF2 and BBLF3"
CDS                  complement(117515. .119080)
                     /note="BBLF2 early reading frame, spliced to BBLF3;
                     Protein sequence is in conflict with the conceptual
                     translation"
                     /codon_start=1
                     /db_xref="PID:e25026"
                     /db_xref="PID:g1334888"
                     /db_xref="SWISS-PROT:P30118"
                     /translation="PDVLKGPVLLRSQTMMETPAESVRARVSSVTFYNVTQTAGRWWA
                     IWVVGIVPIKREDVETLIVVQACQPPLGGSLEPPVVNAPSTTELNFLRWERELRRSGG
                     LIAMLADAAEKDLFDLSFRTRDRRLLSAARVEDEQGLIFQPLFPAQVVCQSCSGDDGR
                     DQQPPPVDGFGSEMEGEQTCPHAQRHSESPGQLDVYIRTPRGDVFTYSTETPDDPSPV
                     PFRDILRPVTYEVDLVSSDGATGRGGDARRHRVSLKILEPAGGFESWLVNSWSMAGGG
                     LYAFLRSIYASCYANHRGTKPIFYLLDPELCPGGSDFQPYVPGFPFLPIHYVGRARPA
                     FWHRAPHSEGLLLLDLNLGVSGTPLADALLGLDARSGQRRGSLLLQQIWPPTRKEINP
                     RHVCTREGGEGGGEDETTVVGRAEATAILEADATWWLYELARCHLSARGAPVGTPDGG
                     GQARDAQTWLRALHRYGTSDTRRALGGLYTAVTRVLLHAAADLGLTWAYADEFILGFV
                     APTSAHPSEEPLAQ"

promoter             118981
                     /note="TATA: TATAAAA BBR1 late promoter before BBRF3"
promoter             119067
                     /note="TATA: TTTAAAA BBR2 late promoter ?"
promoter             119098
                     /note="TATA: TATTTAA BBR3 late promoter before BBRF3"
misc_feature         119108
                     /note="DONOR: AAGGTGAAT possible donor"
CDS                  119137. .120354
                     /note="BBRF3 late reading frame"
                     /codon_start=1
                     /db_xref="PID:g1334889"
                     /db_xref="SWISS-PROT:P03215"
                     /translation="MKSSKNDTFVYRTWVKTLVVYFVMFVMSAVVPITAMFPNLGYPC
                     YFNALVDYGALNLTNYNLAHHLTPTLYLEPPEMFVYITLVFIADCVAFIYYACGEVAL
                     IKARKKVSGLTDLSAWVSAVGSPTVLFLAILKLWSIQVFIQVLSYKHVFLSAFVYFLH
                     FLASVLHACACVTRFSPVWVVKAQDNSIPQDTFLWWVVFYLKPVVTNLYLGCLALETL
                     VFSLSVFLALGNSFYFMVGDMVLGAVNLFLILPIFWYILTEVWLASFLRHNFGFYCGM
                     FIASIILILPLVRYEAVFVSAKLHTTVAINVAIIPILCSVAMLIRICRIFKSMRQGTD
                     YVPVSETVELELESEPRPRPSRTPSPGRNRRRSSTSSSSSRSTRRQRPVSTQALVSSV
                     LPMTTDSEEEIFP"
misc_feature         120260
                     /note="ACCEPT: ATCTTCCTCCAGGT possible acceptor"
misc_feature         120358
                     /note="polyA signal: AATAAA"
CDS                  complement(120747. .120974)
                     /note="BBLF1 late reading frame, possibly homologous to RF
                     49 VZV"
                     /codon_start=1
                     /db_xref="PID:g1334890"
                     /db_xref="SWISS-PROT:P03216"
                     /translation="MGALWSLCRRRVNSIGDVDGGIINLYNDYEEFNLETTKLIAAEE
                     GRACGETNEGLEYDEDSENDELLFLPNKKPN"
misc_feature         complement(120764)
                     /note="polyA signal: AATAAA, 3' end of 0.6kb late, 1.6kb
                     early, 3.0kb early RNAs"
CDS                  complement(120929. .122341)
                     /note="BGLF5 early reading frame, homologous to RF 48 VZV
                     and alkaline exonuclease of HSV"
                     /codon_start=1
                     /db_xref="PID:g1334891"
                     /db_xref="SWISS-PROT:P03217"
                     /translation="MADVDELEDPMEEMTSYTFARFLRSPETEAFVRNLDRPPQMPAM
                     RFVYLYCLCKQIQEFSGETGFCDFVSSLVQENDSKDGPSLKSIYWGLQEATDEQRTVL
                     CSYVESMTRGQSENLMWDILRNGIISSSKLLSTIKNGPTKVFEPAPISTNHYFGGPVA
                     FGLRCEDTVKDIVCKLICGDASANRQFGFMISPTDGIFGVSLDLCVNVESQGDFILFT
                     DRSCIYEIKCRFKYLFSKSEFDPIYPSYTALYKRPCKRSFIRFINSIARPTVEYVPDG
                     RLPSEGDYLLTQDEAWNLKDVRKRKLGPGHDLVADSLAANRGVESMLYVMTDPSENAG
                     RIGIKDRVPVNIFINPRHNYFYQVLLQYKIVGDYVRHSGGGKPGRDCSPRVNIVTAFF
                     RKRSPLDPATCTLGSDLLLDASVEIPVAVLVTPVVLPDSVIRKTLSTAAGSWKAYADN
                     TFDTAPWVPSGLFADDESTP"
promoter             complement(121331)
                     /note="TATA: TATTAAA BBL1 late promoter before BBLF1"
promoter             121669
                     /note="TATA: CATAAAT"
promoter             121697
                     /note="TATA: TATAAAG"
promoter             121772
                     /note="TATA: CATAAAG"
misc_feature         122313
                     /note="BAM: Bam H1 B/G"
CDS                  complement(122328. .123692)
                     /note="BGLF4 early reading frame, homologous to RF 47 VZV;
                     Protein sequence is in conflict with the conceptual
                     translation"
                     /codon_start=1
                     /db_xref="PID:e25030"
                     /db_xref="PID:g1334892"
                     /db_xref="SWISS-PROT:P13288"
                     /translation="SGWRSSVSRSLRPETSCDRPSSHLRNMDVNMAAELSPTNSSSSG
                     ELSVSPEPPRETQAFLGKVTVIDYFTFQHKHLKVTNIDDMTETLYVKLPENMTRCDHL
                     PITCEYLLGRGSYGAVYAHADNATVKLYDSVTELYHELMVCDMIQIGKATAEDGQDKA
                     LVDYLSACTSCHALFMPQFRCSLQDYGHWHDGSIEPLVRGFQGLKDAVYFLNRHCGLF
```

```
                HSDISPSNILVDFTDTMWGMGRLVLTDYGTASLHDRNKMLDVRLKSSKGRQLYRLYCQ
                REPFSIAKDTYKPLCLLSKCYILRGAGHIPDPSACGPVGAQTALRLDLQSLGYSLLYG
                IMHLADSTHKIPYPNPDMGFDRSDPLYFLQFAAPKVVLLEVLSQMWNLNLDMGLTSCG
                ESPCVDVTAEHMSQFLQWCRSLKKRFKESYFFNCRPRFEHPHLPGLVAELLADDFFGP
                DGRRG"
misc_feature    complement(123506)
                /note="DONOR: AAGGTGACT possible donor"
CDS             complement(123941. .124939)
                /note="BGLF3 reading frame"
                /codon_start=1
                /db_xref="PID:g1334894"
                /db_xref="SWISS-PROT:P03220"
                /translation="MFNAVKADMPDDPMLARRYGQCLELALEACQDTPEQFKLVETPL
                KSFLLVSNILPQDNRPWHEARSSGRVAEDDYDFSSLALELLPLNPRLPEEWQFGGQGW
                SSRMEPSQPEMGMGLCFEVFDGDLMRIALAWNKDEVIGQALQILAHSQTWTSLVPEDP
                LPWMWALFYGPRSHCEERHCVYAAARGKRGPILLPTAVYTPCANIEAFLAHLTRCVYA
                LYLDVRDWKGEDIAPPFDVSRLNKMAKQLCLLPQEPFCITRVCLLCLLHKQNLNAQYK
                RPVDTYDPCLILTGEAERYMVDAVGNYREASTGTTVLYPTYDLGSIVADMVTYEDE"
promoter        complement(124117)
                /note="TATA: TATAAAA"
misc_feature    complement(124219)
                /note="polyA signal: AATAAA"
CDS             124938. .125915
                /note="BGRF1 reading frame, homologous to RF 45 VZV and
                spliced HSV gene (Costa et al, 1985). Spliced to BDRF1.
                Northern blots in BGRF1 detect 2.7, 2.6, 2.1kb late and
                1.9kb early RNAs. 2.6, 2.1kb RNAs very weak."
                /codon_start=1
                /db_xref="PID:g1334893"
                /db_xref="SWISS-PROT:P03219"
                /translation="MLYASQRGRLTENLRNALQQDSTTQGCLGAETPSIMYTGAKSDR
                WAHPLVGTIHASNLYCPMLRAYCRHYGPRPVFVASDESLPMFGASPALHTPVQVQMCL
                LPELRDTLQRLLPPPNLEDSEALTEFKTSVSSARAILEDPNFLEMREFVTSLASFLSG
                QYKHKPARLEAFQKQVVLHSFYFLISIKSLEITDTMFDIFQSAFGLEEMTLEKLHIFK
                QKASVFLIPRRHGKTWIVVAIISLILSNLSNVQIGYVAHQKHVASAVFTEIIDTLTKS
                FDSKRVEVNKETSTITFRHSGKISSTVMCATCFNKNVRPDVSVLGNCRA"
CDS             join(124938. .125873,129215. .130351)
                /note="BDRF1 reading frame"
                /codon_start=1
                /product="probable DNA packaging protein"
                /db_xref="PID:e276433"
                /db_xref="PID:g1632798"
                /db_xref="SWISS-PROT:P03219"
                /translation="MLYASQRGRLTENLRNALQQDSTTQGCLGAETPSIMYTGAKSDR
                WAHPLVGTIHASNLYCPMLRAYCRHYGPRPVFVASDESLPMFGASPALHTPVQVQMCL
                LPELRDTLQRLLPPPNLEDSEALTEFKTSVSSARAILEDPNFLEMREFVTSLASFLSG
                QYKHKPARLEAFQKQVVLHSFYFLISIKSLEITDTMFDIFQSAFGLEEMTLEKLHIFK
                QKASVFLIPRRHGKTWIVVAIISLILSNLSNVQIGYVAHQKHVASAVFTEIIDTLTKS
                FDSKRVEVNKETSTITFRHSGKISSTVMCATCFNKNSIRGQTFHLLFVDEANFIKKEA
                LPAILGFMLQKDAKIIFISSVNSADQATSFLYKLKDAQERLLNVVSYVCQEHRQDFDM
                QDSMVSCPCFRLHIPSYITMDSNIRATTNLFLDGAFSTELMGDTSSLSQGSLSRTVRD
                DAINQLELCRVDTLNPRVAGRLASSLYVYVDPAYTNNTSASGTGIAAVTHDRADPNRV
                IVLGLEHFFLKDLTGDAALQIATCVVALVSSIVTLHPHLEEVKVAVEGNSSQDSAVAI
                ASIIGESCPLPCAFVHTKDKTSSLQWPMYLLTNEKSKAFERLIYAVNTASLSASQVTV
                SNTIQLSFDPVLYLISQIRAIKPIPLRDGTYTYTGKQRNLSDDVLVALVMAHFLATTQ
                KHTFKKVH"
promoter        complement(125113)
                /note="TATA: TATAAAT before BGLF3"
misc_feature    complement(125484)
                /note="polyA signal: AATAAA, 3' end of 1.6kb late, 1.8kb
                late, 3.0kb late and 3.7kb early RNAs"
CDS             complement(125863. .126873)
                /note="BGLF2 late reading frame, poor homology to RF 44
                VZV"
                /codon_start=1
                /db_xref="PID:g1334895"
                /db_xref="SWISS-PROT:P03221"
                /translation="MASAANSSREQLRKFLNKECLWVLSDASTPQMKVYTATTAVSAV
                YVPQIAGPPKTYMNVTLIVLKPKKKPTYVTVYINGTLATVARPEVLFTKAVQGPHSLT
                LMYFGVFSDAVGEAVPVEIRGNPVVTCTDLTTAHVFTTSTAVKTVEELQDITPSEIIP
                LGRGGAWYAEGALYMFFVNMDMLMCCPNMPTFPSLTHFINLLTRCDNGECVTCYGAGA
                HVNILRGWTEDDSPGTSGTCPCLLPCTALNNDYVPITGHRALLGLMFKPEDAPFVVGL
                RFNPPKMHPDMSRVLQGVLANGKEVPCTAQPWTLLRFSDLYSRAMLYNCQVLKRQVLH
```

```
                SY"
promoter        126277
                /note="TATA: GATAAAA"
CDS             complement(126851. .128374)
                /note="BGLF1 late reading frame"
                /codon_start=1
                /db_xref="PID:g1334896"
                /db_xref="SWISS-PROT:P03222"
                /translation="MDVHIDNQVLSGLGTPLLVHLFVPDTVMAELCPNRVPNCEGAWC
                QTLFSDRTGLTRVCRVFAARGMLPGRPSHRGTFTSVPVYCDEGLPELYNPFHVAALRF
                YDEGGLVGELQIYYLSLFEGAKRALTDGHLIREASGVQESAAAMQPIPIDPGPPGGAG
                IEHMPVAAAQVEHPKTYDLKQILLEITQEENRGEQRLGHAGSPALCLGLRLRAGAETK
                AAAETSVSKHHPALENPSNIRGSAGGEGGGGRAGTGGTVGVGSGALSRVPVSFSKTRR
                AIRESRALVRGIAHIFSPHALYVVTYPELSAQGRLHRMTAVTHASPATDLAEVSILGA
                PEREFRFLISVALRISASFREKLAMQAWTAQQEIPVVIPTSYSRIYKNSDLIREAFFT
                VQTRVSWESCWVKATISNAPKTPDACLWIDSHPLYEEGASAWGKVIDSRPPGGLVGAA
                SQLVALGTDGHCVHLATTSDGQAFLVLPGGFVIKGQLALTPEERGYILARHGIRREQ"
promoter        complement(126929)
                /note="TATA: TATTAAA EEL8 late promoter before BGLF2,
                gives 1.6kb late RNA"
promoter        complement(127237)
                /note="TATA: TATAAAA, potential promoter for 1.8kb late
                RNA"
misc_feature    128029
                /note="polyA signal: AATAAA"
CDS             complement(128344. .129021)
                /note="BDLF4 early reading frame"
                /codon_start=1
                /db_xref="PID:g1334897"
                /db_xref="SWISS-PROT:P03223"
                /translation="MSDQGRLSLPRGEGGTDEPNPRHLCSYSKLEFHLPLPESMASVF
                ACWGCGEYHVCDGSSECTLIETHEGVVCALTGNYMGPHFQPALRPWTEIRQDTQDQRD
                KWEPEQVQGLVKTVVNHLYHYFLNENVISGVSEALFDQEGALRPHIPALVSFVFPCCL
                MLFRGASSEKVVDVVLSLYIHVIISIYSQKTVYGALLFKSTRNKRYDAVAKRMRELWM
                STLTTKC"
promoter        complement(128432)
                /note="TATA: TATTTAA before BGLF1, potential promoter for
                3.0kb late RNA"
misc_feature    128848
                /note="BAM: Bam H1 G/D"
promoter        complement(129054)
                /note="TATA: TATTTGC before BDLF4, potential promoter for
                3.7kb early RNA"
mat_peptide     129188. .130348
                /note="BDRF1 reading frame, homologous to RF 42 VZV and
                spliced gene in HSV (Costa et al, 1985). Spliced from
                BGRF1. Northern blots in BDRF1 detect 2.7, 2.6 kb late and
                1.9kb early RNAs. Possibly also 1.8kb early RNA."
promoter        129374
                /note="TATA: TATAAGC"
promoter        complement(129377)
                /note="TATA: TATAAAG"
misc_feature    129413
                /note="DONOR: GTGGTAAGT possible donor"
misc_feature    130347
                /note="polyA signal: ATTAAA"
misc_feature    complement(130359)
                /note="polyA signal: AATAAA, 3' end of 0.9kb late RNA,
                2.3kb late RNA and 3.2kb late RNA"
CDS             complement(130362. .131066)
                /note="BDLF3 late reading frame 9xNXT/S"
                /codon_start=1
                /db_xref="PID:g1334899"
                /db_xref="SWISS-PROT:P03224"
                /translation="MAHARDKAGAVMAMILICETSLIWTSSGSSTASAGNVTGTTAVT
                TPSPSASGPSTNQSTTLTTTSAPITTTAILSTNTTTVTFTGTTVTPVPTTSNASTINV
                TTKVTAQNITATEAGTGTSTGVTSNVTTRSSSTTSATTRITNATTLAPTLSSKGTSNA
                TKTTAELPTVPDERQPSLSYGLPLWTLVFVGLTFLMLILIFAAGLMMSAKNKPLDEAL
                LTNAVTRDPSLYKGLV"
promoter        complement(131104)
                /note="TATA: TATAAAA EEL4 late promoter before BDLF3,
                gives 0.9kb late RNA"
CDS             complement(131127. .132389)
```

```
                     /note="BDLF2 late reading frame"
                     /codon_start=1
                     /db_xref="PID:g1334900"
                     /db_xref="SWISS-PROT:P03225"
                     /translation="MVDEQVAVEHGTVSHTISREEDGVVHERRVLASGERVEVFYKAP
                     APRPREGRASTFHDFTVPAAAAVPGPEPEPEPHPPMPIHANGGGETKTNTQDQNQNQT
                     TRTRTNAKAEERTAEMDDTMASSGGQRGAPISADLLSLSSLTGRMAAMAPSWMKSEVC
                     GERMRFKEDVYDGEAETLAEPPRCFMLSFVFIYYCCYLAFLALLAFGFNPLFLPSFMP
                     VGAKVLRGKGRDFGVPLSYGCPTNPFCKVYTLIPAVVINNVTYYPNNTDSHGGHGGFE
                     AAALHVAALFESGCPNLQAVTNRNRTFNVTRASGRVERRLVQDMQRVLASAVVVMHHH
                     CHYETYYVFDGVGPEFGTIPTPCFKDVLAFRPSLVTNCTAPLKTSVKGPNWSGAAGGM
                     KRKQCRVDRLTDRSFPAYLEEVMYVMVQ"
promoter             132266
                     /note="TATA: TATAAAA"
CDS                  complement(132400. .133305)
                     /note="BDLF1 late reading frame, poor homology to RF 41
                     VZV"
                     /codon_start=1
                     /db_xref="PID:g1334901"
                     /db_xref="SWISS-PROT:P25214"
                     /translation="MDLKVVVSLSSRLYTDEIAKMQQRIGCILPLASTHGTQNVQGLG
                     LGQVYSLETVPDYVSMYNYLSDCTLAVLDEVSVDSLILTKIVPGQTYAIKNKYQPFFQ
                     WHGTGSLSVMPPVFGREHATVKLESNDVDIVFPMVLPTPIAEEVLQKILLFNVYSRVV
                     MQAPGNADMLDVHMHLGSVSYLGHHYELALPEVPGPLGLALLDNLSLYFCIMVTLLPR
                     ASMRLVRGLIRHEHHDLLNLFQEMVPDEIARIDLDDLSVADDLSRMRVMMTYLQSLAS
                     LFNLGPRLATAAYSQETLTATCWLR"
promoter             complement(132476)
                     /note="TATA: TATTTAA before BDLF2, likely promoter for
                     2.3kb late RNA"
misc_feature         complement(133312)
                     /note="polyA signal: AATAAA, 3' end of 4.5kb late RNA"
CDS                  complement(133321. .137466)
                     /note="BcLF1 late reading frame, homologous to RF 40 VZV
                     and major capsid protein of HSV"
                     /codon_start=1
                     /db_xref="PID:g1334902"
                     /db_xref="SWISS-PROT:P03226"
                     /translation="MASNEGVENRPFPYLTVDADLLSNLRQSAAEGLFHSFDLLVGKD
                     AREAGIKFEVLLGVYTNAIQYVRFLETALAVSCVNTEFKDLSRMTDGKIQFRISVPTI
                     AHGDGRRPSKQRTFIVVKNCHKHHISTEMELSMLDLEILHSIPETPVEYAEYVGAVKT
                     VASALQFGVDALERGLINTVLSVKLRHAPPMFILQTLADPTFTERGFSKTVKSDLIAM
                     FKRHLLEHSFFLDRAENMGSGFSQYVRSRLSEMVAAVSGESVLKGVSTYTTAKGGEPV
                     GGVFIVTDNVLRQLLTFLGEEADNQIMGPSSYASFVVRGENLVTAVSYGRVMRTFEHF
                     MARIVDSPEKAGSTKSDLPAVAAGVEDQPRVPISAAVIKLGNHAVAVESLQKMYNDTQ
                     SPYPLNRRMQYSYYFPVGLFMPNPKYTTSAAIKMLDNPTQQLPVEAWIVNKNNLLLAF
                     NLQNALKVLCHPRLHTPAHTLNSLNAAPAPRDRRETYSLQHRRPNHMNVLVIVDEFYD
                     NKYAAPVTDIALKCGLPTEDFLHPSNYDLLRLELHPLYDIYIGRDAGERARHRAVHRL
                     MVGNLPTPLAPAAFQEARGQQFETATSLAHVVDQAVIETVQDTAYDTAYPAFFYVVEA
                     MIHGFEEKFVMNVPLVSLCINTYWERSGRLAFVNSFSMIKFICRHLGNNAISKEAYSM
                     YRKIYGELIALEQALMRLAGSDVVGDESVGQYVCALLDPNLLPPVAYTDIFTHLLTVS
                     DRAPQIIIGNEVYADTLAAPQFIERVGNMDEMAAQFVALYGYRVNGDHDHDFRLHLGP
                     YVDEGHADVLEKIFYYVFLPTCTNAHMCGLGVDFQHVAQTLAYNGPAFSHHFTRDEDI
                     LDNLENGTLRDLLEISDLRPTVGMIRDLSASFMTCPTFTRAVRVSVDNDVTQQLAPNP
                     ADKRTEQTVLVNGLVAFAFSERTRAVTQCLFHAIPFHMFYGDPRVAATMHQDVATFVM
                     RNPQQRAVEAFNRPEQLFAEYREWHRSPMGKYAAECLPSLVSISGMTAMHIKMSPMAY
                     IAQAKLKIHPGVAMTVVRTDEILSENILFSSRASTSMFIGTPNVSRREARVDAVTFEV
                     HHEMASIDTGLSYSSTMTPARVAAITTDMGIHTQDFFSVFPAEAFGNQQVNDYIKAKV
                     GAQRNGTLLRDPRTYLAGMTNVNGAPGLCHGQQATCEIIVTPVTADVAYFQKSNSPRG
                     RAACVVSCENYNQEVAEGLIYDHSRPDAAYEYRSTVNPWASQLGSLGDIMYNSSYRQT
                     AVPGLYSPCRAFFNKEELLRNNRGLYNMVNEYSQRLGGHPATSNTEVQFVVIAGTDVF
                     LEQPCSFLQEAFPALSASSRALIDEFMSVKQTHAPIHYGHYIIEEVAPVRRILKFGNK
                     VVF"
misc_feature         complement(133332)
                     /note="DONOR: AAGGTGGTT possible donor"
promoter             complement(133352)
                     /note="TATA: TATTAAA before BDLF1"
promoter             complement(133386)
                     /note="TATA: TATATAA"
misc_feature         135178
                     /note="polyA signal: AATAAA"
promoter             135394
                     /note="TATA: TATAAGT"
misc_feature         136624
                     /note="polyA signal: AATAAA"
misc_feature         136868
                     /note="BAM: Bam H1 D/c"
promoter             complement(137710)
                     /note="TATA: TATTAAA EHL1 promoter before BcLF1, gives
                     4.5kb late RNA"
promoter             137857
                     /note="TATA: CATAAAC"
CDS                  137991. .139718
                     /note="BcRF1 reading frame"
                     /codon_start=1
                     /db_xref="PID:e25010"
                     /db_xref="PID:g1632799"
                     /db_xref="SWISS-PROT:P25215"
                     /translation="MLAHLNQVTRIPPCPPFSGREARLKFHFFSWSTFMLSWPNNATL
                     REIRTRAATNLTHHPHLVDTLYHASPQTPFLTRSGALYRFVTCCNCTLPNISIQQCKA
                     GDRPGDLEIILQSNGGGRPASFQFPSSPTGSLLRCIVAASLLPEVSVGHQELSPLRSR
                     SQGGQTDVRSGPDPARRLVALLRREDGAPKDPPLGPFGHPRGPGPAKSEDEESERRDA
                     PPPPLDSSFQASRLVPVGPGFRLLVFNTNRVINTKLVCSEPLVKMRVCNVPRLINNFV
                     ARKYVVKETAFTVSLFFTDGVGANLAINVNISGTYLSFLLAMTSLRCFLPVEAIYPAA
                     VSNWNSTLDLHGLENQSLVRENRSGVFWTTNFPSVVSCRDGLNVSWFKAATATISRVH
                     GQTLEQHLIREITPIVTHREAKISRIKNRLFTLLELRNRSQIQVLHKRFLEGLLDCAS
                     LLRLDPSCINRIASEGLFDFSKRSIAHSKNRHECALLGHRHSANVTKLVVNERKTRLD
                     ILGRNANFLTRCKHQVNLRQSPIFLTLLRHIRRRLGLGRASVKREITLLLAHLRKKTA
                     PIHCRDAQV"
misc_feature         138019
                     /note="BAM: Bam H1 c/b"
misc_feature         139352
                     /note="BAM: Bam H1 b/T"
CDS                  139642. .140916
                     /note="BTRF1 reading frame. Northern blots detect 0.95
                     late and 3.8kb early RNA; Protein sequence is in conflict
                     with the conceptual translation"
                     /codon_start=1
                     /db_xref="PID:e25011"
                     /db_xref="PID:g1334904"
                     /db_xref="SWISS-PROT:P30119"
                     /translation="NERLPFSWPTCAKRQPPSTAVMLKCKQPGARFIHGAVHLPSGQI
                     VFHTIHSPTLASALGLPGENVPIPALFRASGLNVRESLPMTNMRAPIISLARLILAPN
                     PYILEGQLTVGMTQDNGIPVLFARPVIEVKSGPESNIKASSQLMIAEDSCLNQIAPFS
                     ASEHPAFSMVESVKRVRVDEGANTRRTIRDILEIPVTVLSSLQLSPTKSILKKAPEPP
                     PPEPQATFDATPYARIFYDIGRQVPKLGNAPAAQVSNVLIANRSHNSLRLVPNPDLLP
                     LQHLYLKHVVLKSLNLENIVQDFEAIFTSPSDTISEAETKAFEKLVEQAKNTVENIVF
                     CLNSICSTSTLPDVVPDVNNPNISLALEKYFLMFPPSGTIMRNVRFATPIVRLLCQGA
                     ELGTMAQFLGKYIKVKKETGMYTLVKLYYLLRI"
misc_feature         complement(140902)
                     /note="polyA signal: AATAAA, 3' end of 2.5kb late RNA"
CDS                  complement(140916. .143036)
                     /note="BXLF2 late reading frame, encodes gp85; homologous
                     to RF 37 VZV and glycoprotein H of HSV (gpIII of VZV)"
                     /codon_start=1
                     /db_xref="PID:g1334905"
                     /db_xref="SWISS-PROT:P03231"
                     /translation="MQLLCVFCLVLLWEVGAASLSEVKLHLDIEGHASHYTIPWTELM
                     AKVPGLSPEALWREANVTEDLASMLNRYKLIYKTSGTLGIALAEPVDIPAVSEGSMQV
                     DASKVHPGVISGLNSPACMLSAPLEKQLFYYIGTMLPNTRPHSYVFYQLRCHLSYVAL
                     SINGDKFQYTGAMTSKFLMGTYKRVTEKGDEHVLSLVFGKTKDLPDLRGPFSYPSLTS
                     AQSGDYSLVIVTTFVHYANFHNYFVPNLKDMFSRAVTMTAASYARYVLQKLVLLEMKG
                     GCREPELDTETLTTMFEVSVAFFKVGHAVGETGNGCVDLRWLAKSFFELTVLKDIIGI
                     CYGATVKGMQSYGLERLAAMLMATVKMEELGHLTTEKQEYALRLATVGYPKAGVYSGL
                     IGGATSVLLSAYNRHPLFQPLHTVMRETLFIGSHVVLRELRLNVTTQGPNLALYQLLS
                     TALCSALEIGEVLRGLALGTESGLFSPCYLSLRFDLTRDKLLSMAPQEATLDQAAVSN
                     AVDGFLGRLSLEREDRDAWHLPAYKCVDRLDKVLMIIPLINVTFIISSDREVRGSALY
                     EASTTYLSSSLFLSPVIMNKCSQGAVAGEPRQIPKIQNFTRTQKSCIFCGFALLSYDE
                     KEGLETTTYITSQEVQNSILSSNYFDFDNLHVHYLLLTTNGTVMEIAGLYEERAHVVL
                     AIILYFIAFALGIFLVHKIVMFFL"
misc_feature         140970
                     /note="polyA signal: AATAAA"
misc_feature         complement(141286)
                     /note="polyA signal: AATAAA"
promoter             142589
                     /note="TATA: GATAAAA"
misc_feature         142740
```

```
                /note="BAM: Bam H1 T/X"
CDS             complement(143038. .144861)
                /note="BXLF1 early reading frame, thymidine kinase
                (Littler et al, 1986). Weak homology to RF 36 VZV and HSV
                thymidine kinase. 4.0kb early RNA presumably encodes the
                TK. Also a 2.2kb late RNA here."
                /codon_start=1
                /db_xref="PID:g1334907"
                /db_xref="SWISS-PROT:P03177"
                /translation="MAGFPGKEAGPPGGWRKCQEDESPENERHENFYAEIDDFAPSVL
                TPTGSDSGAGEEDDDGLYQVPTHWPPLMAPTGLSGERVPCRTQAAVTSNTGNSPGSRH
                TSCPFTLPRGAQPPAPAHQKPTAPTPKPRSRECGPSKTPDPFSWFRKTSCTEGGADST
                SRSFMYQKGFEEGLAGLGLDDKSDCESEDESNFRRPSSHSALKQKNGGKGKPSGLFEH
                LAAHGREFSKLSKHAAQLKRLSGSVMNVLNLDDAQDTRQAKAQRKESMRVPIVTHLTN
                HVPVIKPACSLFLEGAPGVGKTTMLNHLKAVFGDLTIVVPEPMRYWTHVYENAIKAMH
                KNVTRARHGREDTSAEVLACQMKFTTPFRVLASRKRSLLVTESGARSVAPLDCWILHD
                RHLLSASVVFPLMLLRSQLLSYSDFIQVLATFTADPGDTIVWMKLNVEENMRRLKKRG
                RKHESGLDAGYLKSVNDAYHAVYCAWLLTQYFAPEDIVKVCAGLTTITTVCHQSHTPI
                IRSGVAEKLYKNSIFSVLKEVIQPFRADAVLLEVCLAFTRTLAYLQFVLVDLSEFQDD
                LPGCWTEIYMQALKNPAIRSQFFDWAGLSKVISDFERGNRD"
promoter        complement(143310)
                /note="TATA: TATAAGA ECL2 late promoter before BXLF2,
                gives 2.5kb late RNA"
misc_feature    143608
                /note="polyA signal: AATAAA"
misc_feature    144791
                /note="ACCEPT: TCTTTCGTTTTCAGG poss. acceptor before
                BXRF1"
CDS             144860. .145606
                /note="BXRF1 late reading frame, homologous to RF 35 VZV.
                Basic (core?) protein."
                /codon_start=1
                /db_xref="PID:g1334906"
                /db_xref="SWISS-PROT:P03232"
                /translation="MDPTRGLCALSTHDLAKFHSLPPARKAAGKRAHLRCYSKLLSLK
                SWEQLASFLSLPPGPTFTDFRLFFEVTLGRRIADCVVVALQPYPRCYIVEFKTAMSNT
                ANPQSVTRKAQRLEGTAQLCDCANFLRTSCPPVLGSQGLEVLAALVFKNQRSLRTLQV
                EFPALGQKTLPTSTTGLLNLLSRWQDGALRARLDRPRPTAQGHRPRTHVGPKPSQLTA
                RVPRSARAGRAGGRKGQVGAVGQVCPGAQK"
misc_feature    144862
                /note="BAM: Bam H1 X/V"
misc_feature    144945
                /note="DONOR: CAGGTAAGC possible donor at 3' BXRF1"
promoter        complement(145135)
                /note="TATA: TATAACA before BXLF1"
promoter        145302
                /note="TATA: TATTTAA before BVRF1, potential promoter for
                1.9kb early RNA"
CDS             145416. .147128
                /note="BVRF1 early reading frame, homologous to RF 34 VZV"
                /codon_start=1
                /db_xref="PID:g1334908"
                /db_xref="SWISS-PROT:P03233"
                /translation="MALSGHVLIDPARLPRDTGPELMWAPSLRNSLRVSPEALELAER
                EAERARSERWDRCAQVLKNRLLRVELDGIMRDHLARAEEIRQDLDAVVAFSDGLESMQ
                VRSPSTGGRSAPAPPSPSPAQPFTRLTGNAQYAVSISPTDPPLMVAGSLAQTLLGNLY
                GNINQWVPSFGPWYRTMSANAMQRRVFPKQLRGNLNFTNSVSLKLMTEVVAVLEGTTQ
                DFFSDVRHLPDLQAALILSVAYLLLQGGSSHQQRPLPASREELLELGPESLEKIIADL
                KAKSPGGNFMILTSGNKEARQSIAPLNRQAAYPPGTFADNKIYNLFVGAGLLPTTAAL
                NVPGAAGRDRDLVYRIANQIFGEDVPPFSSHQWNLRVGLAALEALMLVYTLCETANLA
                EAATRRLHLSSLLPQAMQRRKPAMASAGMPGAYPVQTLFRHGELFRFIWAHYVRPTVA
                ADPQASISSLFPGLVLLALELKLMDGQAPSHYAINLTGQKFDTLFEIINQKLLFHDPA
                AMLAARTQLRLAFEDGVGVALGRPSPMLAAREILERQFSASDDYDRLYFLTLGYLASP
                VAPS"
misc_feature    complement(146926)
                /note="polyA signal: AATAAA"
misc_feature    147167
                /note="DONOR: AAGGTAAAT possible donor"
misc_feature    147170
                /note="polyA signal: AATAAA, 3' end of 2.4kb late and
                1.9kb early RNAs"
promoter        147721
                /note="TATA: TATTTAT before BVRF2, potential promoter for
                2.1kb early RNA"
CDS             147927. .149744
                /note="BVRF2 early reading frame, N-terminus homologous to
                RF 33 VZV"
                /codon_start=1
                /db_xref="PID:g1334909"
                /db_xref="SWISS-PROT:P03234"
                /translation="MVQAPSVYVCGFVERPDAPPKDACLHLDPLTVKSQLPLKKPLPL
                TVEHLPDAPVGSVFGLYQSRAGLFSAASITSGDFLSLLDSIYHDCDIAQSQRLPLPRE
                PKVEALHAWLPSLSLASLHPDIPQTTADGGKLSFFDHVSICALGRRRGTTAVYGTDLA
                WVLKHFSDLEPSIAAQIENDANAAKRESGCPEDHPLPLTKLIAKAIDAGFLRNRVETL
                RQDRGVANIPAESYLKASDAPDLQKPDKALQSPPPASTDPATMLSGNAGEGATACGGS
                AAAGQDLISVPRNTFMTLLQTNLDNKPPRQTPLPYAAPLPPFSHQAIATAPSYGPGAG
                AVAPAGGYFTSPGGYYAGPAGGDPGAFLAMDAHTYHPHPHPPPAYFGLPGLFGPPPPV
                PPYYGSHLRADYVPAPSRSNKRKRDPEEDEEGGGLFPGEDATLYRKDIAGLSKSVNEL
                QHTLQALRRETLSYGHTGVGYCPQQGPCYTHSGPYGFQPHQSYEVPRYVPHPPPPPTS
                HQAAQAQPPPPGTQAPEAHCVAESTIPEAGAAGNSGPREDTNPQQPTTEGHHRGKKLV
                QASASGVAQSKEPTTPKAKSVSAHLKSIFCEELLNKRVA"
misc_feature    148007
                /note="BAM: Bam H1 V/d"
promoter        148620
                /note="TATA: TATTTAA ECR1 late promoter before BdRF1,
                gives 1.2kb late RNA"
CDS             148707. .149744
                /note="BdRF1 reading frame; this is the C terminus of
                BVRF2"
                /codon_start=1
                /db_xref="PID:g1334910"
                /db_xref="SPTREMBL:Q66541"
                /translation="MLSGNAGEGATACGGSAAAGQDLISVPRNTFMTLLQTNLDNKPP
                RQTPLPYAAPLPPFSHQAIATAPSYGPGAGAVAPAGGYFTSPGGYYAGPAGGDPGAFL
                AMDAHTYHPHPHPPPAYFGLPGLFGPPPPVPPYYGSHLRADYVPAPSRSNKRKRDPEE
                DEEGGGLFPGEDATLYRKDIAGLSKSVNELQHTLQALRRETLSYGHTGVGYCPQQGPC
                YTHSGPYGFQPHQSYEVPRYVPHPPPPPTSHQAAQAQPPPPGTQAPEAHCVAESTIPE
                AGAAGNSGPREDTNPQQPTTEGHHRGKKLVQASASGVAQSKEPTTPKAKSVSAHLKSI
                FCEELLNKRVA"
misc_feature    149115
                /note="BAM: Bam H1 d/I"
misc_feature    149727
                /note="polyA signal: AATAAA, 3' end of 2.1kb early and
                1.2kb late RNAs"
misc_feature    complement(149758)
                /note="polyA signal: AATAAA, 3' end of 1.0kb late, 1.5kb
                late and 1.8kb late RNAs"
CDS             complement(149779. .150525)
                /note="BILF2 late reading frame 11xNXT/S"
                /codon_start=1
                /db_xref="PID:g1334911"
                /db_xref="SWISS-PROT:P03218"
                /translation="MTHLVLLLCCCVGSVCAFFSDLVKFENVTAHAGARVNLTCSVPS
                NESVSRIELGRGYTPGDGQLPLAVATSNNGTHITNGGYNYSLTLEWVNDSNTSVSLII
                PNVTLAHAGYYTCNVTLRNCSVASGVHCNYSAGEEDDQYHANRTLTQRMHLTVIPATT
                IAPTTLVSHTTSTSHRPHRRPVSKRPTHKPVTLGPFPIDPWRPKTTWVHWALLLITCA
                VVAPVLLIIIISCLGWLAGWGRRRKGWIPL"
promoter        complement(150571)
                /note="TATA: TATTTAG before BILF2. Potential promoter for
                1.0kb late RNA."
repeat_region   151236. .151618
                /note="repetitive sequence 3X"
misc_feature    151767
                /note="polyA signal: AATAAA"
promoter        complement(151780)
                /note="TATA: CATAAAA"
misc_feature    152012. .152013
                /note="DEL: B95-8 deletion with respect to Raji"
CDS             complement(152161. .153099)
                /note="BILF1 reading frame, membrane protein, 3xNXS /T"
                /codon_start=1
                /db_xref="PID:g1334912"
                /db_xref="SWISS-PROT:P03208"
                /translation="MLSTMAPGSTVGTLVANMTSVNATEDACTKSYSAFLSGMTSLLL
                VLLILLTLAGILFIIFVRKLVHRMDVWLIALLIELLLWVLGKMIQEFSSTGLCLLTQN
                MMFLGLMCSVWTHLGMALEKTLALFSRTPKRTSHRNVCLYLMGVFCLVLLLIIILLIT
```

MGPDANLNRGPNMCREGPTKGMHTAVQGLKAGCYLLAAVLIVLLTVIIIWKLLRTKFG
RKPRLICNVTFTGLICAFSWFMLSLPLLFLGEAGSLGFDCTESLVARYYPGPAACLAL
LLIILYAWSFSHFMDSLKNQVTVTARYFRRVPSQST"

promoter 152230
/note="TATA: CATAAAA"

misc_feature 153259
/note="polyA signal: AATAAA"

misc_feature 153637
/note="HPN: 22bp 2-fold symmetric"

misc_feature complement(153690)
/note="DONOR: AAAGTGAGG possible donor"

CDS complement(153699. .156746)
/note="BALF5 DNA polymerase (early), homologous to many DNA polymerases, CMV HFLF2 and RF 28 VZV. 4.5kb early RNA apparently encodes BALF5, RNA ends unknown."
/codon_start=1
/db_xref="PID:g1334913"
/db_xref="SWISS-PROT:P03198"
/translation="MSGGLFYNPFLRPNKGLLKKPDKEYLRLIPKCFQTPGAAGVVDV
RGPQPPLCFYQDSLTVVGGDEDGKGMWWRQRAQEGTARPEADTHGSPLDFHVYDILET
VYTHEKCAVIPSDKQGYVVPCGIVIKLLGRRKADGASVCVNVFGQQAYFYASAPQGLD
VEFAVLSALKASTFDRRTPCRVSVEKVTRRSIMGYGNHAGDYHKITLSHPNSVCHVAT
WLQDKHGCRIFEANVDATRRFVLDNDFVTFGWYSCRRAIPRLQHRDSYAELEYDCEVG
DLSVRREDSSWPSYQALAFDIECLGEEGFPTATNEADLILQISCVLWSTGEEAGRYRR
ILLTLGTCEDIEGVEVYEFPSELDMLYAFFQLIRDLSVEIVTGYNVANFDWPYILDRA
RHIYSINPASLGKIRAGGVCEVRRPHDAGKGFLRANTKVRITGLIPIDMYAVCRDKLS
LSDYKLDTVARHLLGAKKEDVHYKEIPRLFAAGPEGRRRLGMYCVQDSALVMDLLNHF
VIHVEVAEIAKIAHIPCRRVLDDGQQIRVFSCLLAAAQKENFILPMPSASDRDGYQGA
TVIQPLSGFYNSPVLVVDFASLYPSIIQAHNLCYSTMITPGEEHRLAGLRPGEDYESF
RLTGGVYHFVKKHVHESFLASLLTSWLAKRKAIKKLLAACEDPRQRTILDKQQLAIKC
TCNAVYGFTGVANGLFPCLSIAETVTLQGRTMLERAKAFVEALSPANLQALAPSPDAW
APLNPEGQLRVIYGDTDSLFIECRGFSESETLRFADALAAHTTRSLFVAPISLEAEKT
FSCLMLITKKRYVGVLTDGKTLMKGVELVRKTACKFVQTRCRRVLDLVLADARVKEAA
SLLSHRPFQESFTQGLPVGFLPVIDILNQAYTDLREGRVPMGELCFSTELSRKLSAYK
STQMPHLAVYQKFVERNEELPQIHDRIQYVFVEPKGGVKGARKTEMAEDPAYAERHGV
PVAVDHYFDKLLQGAANILQCLFDNNSGAALSVLQNFTARPPF"

misc_feature 154747
/note="BAM: Bam H1 I/A"

misc_feature complement(156707)
/note="polyA signal: AATAAA; 3' end of 2.5kb late (gB) RNA and 1.8kb late RNA"

CDS complement(156749. .159322)
/note="BALF4 late reading frame 9xNXT/S homologous to HSV1 glycoprotein B (Pellet et al, 1985), CMV HFLF1 and RF 31 VZV (gpII)"
/codon_start=1
/db_xref="PID:g1334914"
/db_xref="SWISS-PROT:P03188"
/translation="MTRRRVLSVVVLLAALACRLGAQTPEQPAPPATTVQPTATRQQT
SFPFRVCELSSHGDLFRFSSDIQCPSFGTRENHTEGLLMVFKDNIIPYSFKVRSYTKI
VTNILIYNGWYADSVTNRHEEKFSVDSYETDQMDTIYQCYNAVKMTKDGLTRVYVDRD
GVNITVNLKPTGGLANGVRRYASQTELYDAPGWLIWTYRTRTTVNCLITDMMAKSNSP
FDFFVTTTGQTVEMSPFYDGKNKETFHERADSFHVRTNYKIVDYDNRGTNPQGERRAF
LDKGTYTLSWKLENRTAYCPLQHWQTFDSTIATETGKSIHFVTDEGTSSFVTNTTVGI
ELPDAFKCIEEQVNKTMHEKYEAVQDRYTKGQEAITYFITSGGLLLAWLPLTPRSLAT
VKNLTELTTPTSSPPSSPSPPAPSAARGSTPAAVLRRRRRDAGNATTPVPPTAPGKSL
GTLNNPATVQIQFAYDSLRRQINRMLGDLARAWCLEQKRQNMVLRELTKINPTTVMSS
IYGKAVAAKRLGDVISVSQCVPVNQATVTLRKSMRVPGSETMCYSRPLVSFSFINDTK
TYEGQLGTDNEIFLTKKMTEVCQATSQYYFQSGNEIHVYNDYHHFKTIELDGIATLQT
FISLNTSLIENIDFASLELYSRDEQRASNVFDLEGIFREYNFQAQNIAGLRKDLDNAV
SNGRNQFVDGLGELMDSLGSVGQSITNLVSTVGGLFSSLVSGFISFFKNPFGGMLILV
LVAGVVILVISLTRRTRQMSQQPVQMLYPGIDELAQQHASGEGPGINPISKTELQAIM
LALHEQNQEQKRAAQRAAGPSVASRALQAARDRFPGLRRRRYHDPETAAALLGEAETE
F"

promoter 158204
/note="TATA: TATAAAA"

CDS complement(159312. .161678)
/note="BALF3 reading frame; Protein sequence is in conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e24996"
/db_xref="PID:g1334915"
/db_xref="SWISS-PROT:P25939"
/translation="GRRGVLIGPLLRPGGQRPRNPGDHCLQRDRVDGGGRSGLSGGSV
GWARGSPGQLYQRGRAGQHRVAGLGRAAAGWRRAGLRGQTQAPSGHRSPRTRGLETPG
AAMSGLLAAAYSQVYALAVELSVCTRLDPRSLDVAAVVRNAGLLAELEAILLPRLRRQ
NDRACSALSLELVHLLENSREASAALLAPGRKGTRVPPLRTPSVAYSVEFYGGHKVDV
SLCLINDIEILMKRINSVFYCMSHTMGLESLERALDLLGRFRGVSPIPDPRLYITSVP
CWRCVGELMVLPNHGNPSTAEGTHVSCNHLAVPVNPEPVSGLFENEVRQAGLGHLLEA
EEKARPGGPEEGAVPGPGRPEAEGATRALDTYNVFSTVPPEVAELSELLYWNSGGHAI
GATGQGEGGGHSRLSALFARERRLALVRGACEEALAGARLTHLFDAVAPGATERLFCG
GVYSSSGDAVEALKADCAAAFTAHPQYRAILQKRNELYTRLNRAMQRLGRGEEEASRE
SPEVPRPAGAREPGPSGALSDALKRKEQYLRQVATEGLAKLQSCLAQQSETLTETLCL
RVWGDVVYWELARMRNHFLYRRAFVSGPWEDRRAGEGAAFENSKYIKTHLFTQTLSSE
HLHALTHSLYTFITGPLAEESGLFPPPSNVALARCCDAAGTLPHQKAFLTSLIWPGIE
PSDWIETSFNSFYSVPGGSLASSQQILCRALREAVLTVSLYNKTWGRSLILRRADAVS
PGQALPPDGLYLTYDSDRPLILLYKGRGWVFKDLYALLYLHLQMRDDSA"

promoter complement(159370)
/note="TATA: TATTTAA ECL1 late promoter before BALF4, gives 2.5kb late RNA"

misc_feature 160966
/note="polyA signal: AATAAA"

misc_feature 160990
/note="polyA signal: ATTAAA"

misc_feature complement(161013)
/note="polyA signal: AATAAA, presumed end of 3.9kb early RNA"

CDS complement(161384. .164770)
/note="BALF2 early reading frame, homologous to RF 29 VZV and major DNA binding protein HSV. 3.9kb early RNA"
/codon_start=1
/db_xref="PID:g1334916"
/db_xref="SWISS-PROT:P03227"
/translation="MQGAQTSEDNLGSQSQPGPCGYIYFYPLATYPLREVATLGTGYA
GHRCLTVPLLCGITVEPGFSINVKALHRRPDPNCGLLRATSYHRDIYVFHNAHMVPPI
FEGPGLEALCGETREVFGYDAYSALPRESSKPGDFFPEGLDPSAYLGAVAITEAFKER
LYSGNLVAIPSLKQEVAVGQSASVRVPLYDKEVFPEGVPQLRQFYNSDLSRCMHEALY
TGLAQALRVRRVGKLVELLEKQSLQDQAKVAKVAPLKEFPASTISHPDSGALMIVDSA
ACELAVSYAPAMLEASHETPASLNYDSWPLFADCEGPEARVAALHRYNASLAPHVSTQ
IFATNSVLYVSGVSKSTGQGKESLFNSFYMTHGLGTLQEGTWDPCRRPCFSGWGGPDV
TGTNGPGNYAVEHLVYAASFSPNLLARYAYYLQFCQGQKSSLTPVPETGSYVAGAAAS
PMCSLCEGRAPAVCLNTLFFRLRDRFPPVMSTQRRDPYVISGASGSYNETDFLGNFLN
FIDKEDDGQRPDDEPRYTYWQLNQNLLERLSRLGIDAEGKLEKEPHGPRDFVKMFKDV
DAAVDAEVVQFMNSMAKNNITYKDLVKSCYHVMQYSCNPFAQPACPIFTQLFYRSLLT
ILQDISLPICMCYENDNPGLGQSPPEWLKGHYQTLCTNFRSLAIDKGVLTAKEAKVVH
GEPTCDLPDLDAALQGRVYGRRLPVRMSKVLMLCPRNIKIKNRVVFTGENAALQNSFI
KSTTRRENYIINGPYMKFLNTYHKTLFPDTKLSSLYLWHNFSRRRSVPVPSGASAEEY
SDLALFVDGGSRAHEESNVIDVVPGNLVTYAKQRLNNAILKACGQTQFYISLIQGLVP
RTQSVPARDYPHVLGTRAVESAAAYAEATSSLTATTVVCAATDCLSQVCKARPVVTLP
VTINKYTGVNGNNQIFQAGNLGYFMGRGVDRNLLQAPGAGLRKQAGGSSMRKKFVFAT
PTLGLTVKRRTQAATTYEIENIRAGLEAIISQKQEEDCVFDVVCNLVDAMGEACASLT
RDDAEYLLGRFSVLADSVLETLATIASSGIEWTAEAARDFLEGVWGGPGAAQDNFISV
AEPVSTASQASAGLLLGGGGQGSGGRRKRRLATVLPGLEV"

misc_feature 163978. .166635
/note="DEL: deletion in Raji"

promoter complement(164814)
/note="TATA: CATTTAA before BALF2, presumed promoter for 3.9kb early RNA"

misc_feature complement(164851)
/note="polyA signal: AATAAA"

CDS complement(164855. .165517)
/note="BALF1 early reading frame, 0.7kb early RNA"
/codon_start=1
/db_xref="PID:g1334918"
/db_xref="SWISS-PROT:P03229"
/translation="MNLAIALDSPHPGLASYTILPRPFYHISLKPVSWPDETMRPAKS
TDSVFVRTPVEAWVAPSPPDDKVAESSYLMFRAMYAVFTRDEKDLPLPALVLCRLIKA
SLRKDRKLYAELACRTADIGGKDTHVRLIISVLRAVYNDHYDYWSRLRVVLCYTVVFA
VRNYLDDHKSAAFVLGAIAHYLALYRRLWFARLGGMPRSLRRQFPVTWALASLTDFLK
SL"

promoter 165442
/note="TATA: GATAAAA"

promoter 165466
/note="TATA: TATAAGA EDR1 early promoter before BARF1, gives 0.8kb early RNA"

CDS 165504. .166169

```
                     /note="BARF1 reading frame"
                     /codon_start=1
                     /db_xref="PID:g1334917"
                     /db_xref="SWISS-PROT:P03228"
                     /translation="MARFIAQLLLLASCVAAGQAVTAFLGERVTLTSYWRRVSLGPEI
                     EVSWFKLGPGEEQVLIGRMHHDVIFIEWPFRGFFDIHRSANTFFLVVTAANISHDGNY
                     LCRMKLGETEVTKQEHLSVVKPLTLSVHSERSQFPDFSVLTVTCTVNAFPHPHVQWLM
                     PEGVEPAPTAANGGVMKEKDGSLSVAVDLSLPKPWHLPVTCVGKNDKEEAHGVYVSGY
                     LSQ"
promoter             complement(165713)
                     /note="TATA: TATAAAG before BALF1"
misc_feature         166165
                     /note="polyA signal: AATAAA 3' end of 0.8kb early RNA.
                     Also 1kb late RNA in this region."
promoter             166469. .166475
                     /note="TATA: TTATTTT"
mRNA                 166498. .166916
                     /note="Exon 1 of terminal protein RNA"
misc_feature         166561. .166563
                     /note="Likely initiator met of terminal protein"
misc_feature         166614
                     /note="BAM: Bam H1 A/Nhet"
misc_feature         complement(166946)
                     /note="polyA signal: AATAAA These 2 polyA sites are 3' end
                     of 0.8kb early, 2.5kb late and 2.5kb latent RNAs"
misc_feature         complement(166950)
                     /note="polyA signal: AATAAA"
CDS                  complement(166998. .167303)
                     /note="BNLF2b reading frame; Protein sequence is in
                     conflict with the conceptual translation"
                     /codon_start=1
                     /db_xref="PID:e25001"
                     /db_xref="PID:g1334919"
                     /db_xref="SPTREMBL:Q04362"
                     /translation="EATMRPGRPLAGFYATLRRSFRRMSKRSKNKAKKERVPVEDRPP
                     TPMPTSQRLIRRNALGGGVRPDAEDCIQRFHPLEPALGVSTKNFDLLSLRCELGWCG"
CDS                  complement(167304. .167486)
                     /note="BNLF2a reading frame"
                     /codon_start=1
                     /db_xref="PID:g1334920"
                     /db_xref="SPTREMBL:Q04361"
                     /translation="MVHVLERALLEQQSSACGLPGSSTETRPSHPCPEDPDVSRLRLL
                     LVVLCVLFGLLCLLLI"
promoter             167320
                     /note="TATA: CATAAAA"
promoter             complement(167525)
                     /note="TATA: TATAAAA EDL2 early promoter before BNLF2a,b.
                     Gives 0.8kb RNA"
mRNA                 complement(168163. .168965)
                     /note="BNLF1 coding part of exon c of latent membrane
                     protein"
repeat_region        168399. .168574
                     /note="5 x 33bp repeats"
intron               complement(168966. .169041)
                     /note="intervening sequence in BNLF1"
misc_feature         complement(168966)
                     /note="ACCEPT: TTTTTTCAACGCAGT possible acceptor"
misc_feature         complement(169041)
                     /note="DONOR: TTGGTAAGA donor"
mRNA                 complement(169042. .169128)
                     /note="BNLF1 exon b of latent membrane protein mRNA"
intron               complement(169129. .169206)
                     /note="intervening sequence in BNLF1"
misc_feature         complement(169129)
                     /note="ACCEPT: TCCTCTTTCCCCAGT acceptor"
promoter             complement(169201)
                     /note="TATA: TATTACA EDL1A late promoter, gives 2.5kb late
                     RNA"
misc_feature         complement(169206)
                     /note="DONOR: TGAGTAAGT donor"
mRNA                 complement(169207. .169474)
                     /note="BNLF1 exon a of latent membrane protein mRNA"
promoter             complement(169546)
                     /note="TATA: TACATAAGC EDL1 promoter before BNLF1 gives
                     2.5kb latent RNA (LMP)"
repeat_region        170094. .170631
                     /note="terminal repeat 1 538bp"
repeat_region        170632. .171154
                     /note="terminal repeat 2 523bp"
repeat_region        171155. .171692
                     /note="terminal repeat 3 538bp"
repeat_region        171693. .172231
                     /note="terminal repeat 4 538bp"
BASE COUNT    34054 a  52511 c  50755 g  34961 t
ORIGIN
```

```
  Query Match             5.6%;  Score 158.8;  DB 17;  Length 172281;
  Best Local Similarity   53.3%;  Pred. No. 1.4e-17;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy    1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
           ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db  108932 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 108873

Qy    1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
            || |  | |       | ||| |  | || || ||  ||||  | || | ||| ||| |
Db  108872 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 108813

Qy    1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
           | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db  108812 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 108753

Qy    1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
              ||||   | ||  |  ||| | |      | ||| ||||| || || |  | ||||||
Db  108752 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 108693

Qy    1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
            || || ||| ||||      || · | ||    |  | || ||| | ||| |  || |
Db  108692 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 108633

Qy    1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
           |||||||  | ||  | ||  | ||  |  |  || |||||    |          |  ||
Db  108632 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 108573

Qy    1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
            ||| | |  |  | ||| |  || | ||  || | ||  | |||||||||     || |
Db  108572 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 108513

Qy    1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
              |||  ||  || || |||||||| || | ||  |  |  ||||     |  | |
Db  108512 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 108453

Qy    1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
            || |  | ||  | ||  | ||| || |  | || ||| |  ||    |  | |   ||
Db  108452 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 108393

Qy    2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
           |  || ||  ||||| |  || || | || || ||| ||    | |||||  ||   | |
Db  108392 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 108333

Qy    2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
            |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db  108332 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 108273

Qy    2153 GACCCCCCTCTCCC 2166
              || |||   ||
Db  108272 CCTCCTCCTGCTCC 108259
```

```
RESULT  12
HS4B958RAJ/c
LOCUS       HS4B958RAJ 184113 bp    DNA             VRL       12-APR-1996
DEFINITION  Epstein-Barr virus, artifactual joining of B95-8 complete genome
            and the sequences from Raji of the large deletion found in B95-8.
ACCESSION   M80517 M75989
```

```
NID         g330330
KEYWORDS    .
SOURCE      Human herpesvirus 4 DNA.
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (sites)
  AUTHORS   Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J.,
            Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C.,
            Tuffnell,P.S. and Barrell,B.G.
  TITLE     DNA sequence and expression of the B95-8 Epstein-Barr virus genome
  JOURNAL   Nature 310 (5974), 207-211 (1984)
  MEDLINE   84270667
REFERENCE   2  (sites)
  AUTHORS   Parker,B.D., Bankier,A., Satchwell,S., Barrell,B. and Farrell,P.J.
  TITLE     Sequence and transcription of Raji Epstein-Barr virus DNA spanning
            the B95-8 deletion region
  JOURNAL   Virology 179 (1), 339-346 (1990)
  MEDLINE   91021036
REFERENCE   3  (sites)
  AUTHORS   Sample,J., Brooks,L., Sample,C., Young,L., Rowe,M., Gregory,C.,
            Rickinson,A. and Kieff,E.
  TITLE     Restricted Epstein-Barr virus protein expression in Burkitt
            lymphoma is due to a different Epstein-Barr nuclear antigen 1
            transcriptional initiation site
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88 (14), 6343-6347 (1991)
  MEDLINE   91296817
REFERENCE   4  (bases 1 to 184113)
  AUTHORS   Jenson,H.B.
  TITLE     GenBank Curator Program
  JOURNAL   Unpublished (1992)
COMMENT     The B95-8 genome (V01555) has a large deletion in the right side of
            the genome which has been sequenced in Raji (M35547).  These
            sequences have been joined to form an extended and more complete,
            although artifactual, EBV sequence.
            For features, refer to feature tables of V01555 and M35547.
FEATURES             Location/Qualifiers
     source          1. .184113
                     /organism="Human herpesvirus 4"
                     /db_xref="taxon:10376"
     misc_feature    1. .152008
                     /note="B95-8 sequences (corresponds to 1-152,008 of
                     V01555)"
     misc_feature    152009. .152012
                     /note="Overlap of B95-8 and Raji sequences at B95-8
                     deletion point (corresponds to 152,009-152,012 in V01555,
                     and 1-4 in M35547)"
     misc_feature    153013. .163839
                     /note="Raji sequences (corresponds to 5-11,831 of M35547)"
     misc_feature    163840. .163843
                     /note="Overlap of B95-8 and Raji sequences at B95-8
                     deletion point (corresponds to 152,009-152,012 of V01555,
                     and 11,832-11,835 of M35547)"
     misc_feature    163844. .184113
                     /note="B95-8 sequences (corresponds to 152,013-172,282 of
                     V01555)"
BASE COUNT    36002 a  55824 c  54622 g  37665 t
ORIGIN


  Query Match             5.6%;  Score 158.8;  DB 17;  Length 184113;
  Best Local Similarity   53.3%;  Pred. No. 1.4e-17;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
           ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db 108932 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 108873

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
            || |  | |      | ||| |  | || || ||  ||||  | || | ||| ||| |
Db 108872 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 108813

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
           | ||   |||| | ||  | ||  |  |||||||  | ||  |||||  || || |||||
Db 108812 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 108753

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
              ||||   | ||  |  ||| | |      | ||| |||||  || || |  |  ||||||
Db 108752 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 108693

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
           || || ||| ||||      ||   | ||    |  | || ||| | ||| |  || |
Db 108692 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 108633

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
           |||||||  | ||  | ||  | ||  |  |  || |||||    |        |  ||
Db 108632 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 108573

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
           ||| | |  |  | ||| |  || | ||  || | ||  | ||||||||     || |
Db 108572 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 108513

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
             |||  ||  || || |||||||| || | ||  |  |  ||||     |  | |
Db 108512 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 108453

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
           || |  | ||  | ||  | ||| || |  | || |||  |  ||    |  | |   ||
Db 108452 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 108393

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
          |  || ||  |||| |  || || | || || ||| ||    | |||| ||   | |
Db 108392 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 108333

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
          |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db 108332 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 108273

Qy   2153 GACCCCCCTCTCCC 2166
             || |||   ||
Db 108272 CCTCCTCCTGCTCC 108259


RESULT  13
HS4ULIR3/c
LOCUS       HS4ULIR3     1150 bp    DNA             VRL       15-JUN-1990
DEFINITION  epstein-barr virus simple repeat array (ir3).
ACCESSION   J02079
NID         g330445
KEYWORDS    repeat region.
SOURCE      ebv (epstein barr virus) from human.
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 1150)
  AUTHORS   Heller,M., van Santen,V.L. and Keiff,E.
  TITLE     simple repeat sequence in epstein-barr virus dna is transcribed in
            latent and productive infections
  JOURNAL   J. Virol. 44, 311-320 (1982)
  MEDLINE   83059881
FEATURES             Location/Qualifiers
     source          1. .1150
                     /organism="Human herpesvirus 4"
                     /db_xref="taxon:10376"
BASE COUNT      302 a    169 c    633 g     46 t
ORIGIN


  Query Match             5.6%;  Score 158.8;  DB 17;  Length 1150;
  Best Local Similarity   53.3%;  Pred. No. 2.2e-17;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
           ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db    987 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 928
```

```
Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
Db    927 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 868

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db    867 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 808

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db    807 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 748

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db    747 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 688

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
Db    687 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 628

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
Db    627 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 568

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
Db    567 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 508

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
Db    507 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 448

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
Db    447 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 388

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
Db    387 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 328

Qy   2153 GACCCCCCTCTCCC 2166
Db    327 CCTCCTCCTGCTCC 314
```

```
RESULT  14
D88734
LOCUS       D88734       3173 bp    DNA             VRL       19-MAY-1997
DEFINITION  Equine herpesvirus 1 DNA for membrane glycoprotein, complete cds.
ACCESSION   D88734
NID         g2114322
KEYWORDS    membrane glycoprotein.
SOURCE      Equine herpesvirus 1 (strain:BK343, isolate:3F clone) DNA.
  ORGANISM  Equine herpesvirus 1
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 3173)
  AUTHORS   Kirisawa,R., Kobayashi,T., Kawakami,Y. and Iwai,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-1996) to the DDBJ/EMBL/GenBank databases. Rikio
            Kirisawa, Rakuno Gakuen University, Faculty of Veterinary Medicine,
            Department of Veterinary Microbiology; Bunkyodai-Midorimachi 582,
            Ebetsu, Hokkaido 069, Japan (E-mail:r-kirisa@ca2.so-net.or.jp,
            Tel:+81-11-386-1112, Fax:+81-11-387-5890)
REFERENCE   2  (sites)
  AUTHORS   Kirisawa,R., Kobayashi,T., Kawakami,Y. and Iwai,H.
  TITLE     Nucleotide sequences of open reading frames 1, 24 and 71 of an
            attenuated equine herpesvirus-1
  JOURNAL   J. Equine Sci. 7, 79-87 (1996)
FEATURES             Location/Qualifiers
     source          1. .3173
                     /organism="Equine herpesvirus 1"
                     /isolate="3F clone"
                     /strain="BK343"
                     /db_xref="taxon:10326"
     gene            118. .2721
                     /gene="ORF71"
     CDS             118. .2721
                     /gene="ORF71"
                     /codon_start=1
                     /product="membrane glycoprotein"
                     /db_xref="PID:d1020846"
                     /db_xref="PID:g2114323"
                     /translation="MGFIYARKLLLCMAVSIYAIGSTTTTETTTSSSSTSGSGQSTSS
                     GTTNSSSSPTTSPPTTSSSPPTSTHTSSPSSTSTQSSSTAATSSSAPSTASSTTSIPT
                     STSTETTTTTPTASTTTPTTTTAAPTTAATTTAVTTAASTSAETTTATATATSTPTTT
                     TPTSTTTTTATTTVPTTASTTTDTTTAATTTAATTTAATTTAATTTAATTTAATTTAA
                     TTTAATTTAATTTAATTTAATTTAATTTAATTTAATTSSATTAATTTAATTTAATTTA
                     ATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTGSPT
                     SGSTSTTGASTSTPSASTATSATPTSTSTSAAATTSTPTPTSAATSAESTTEAPTSTP
                     TTDTTTPSEATTATTSPESTTVSASTTSATTTAFTTESHTSPDSSTGSTSTAEPSSTF
                     TLTPSTATPSTDQFTGSSASTESDSTDSSTVPTTGTESITESSSTTEASTNLGSSTYE
                     STEALETPDGNTTSGNTTPSPSPRTPSFADTQQTPDNGVSTQHTTINDHTTANAQKHA
                     GYHRGRAGGRRGIPQGGSHTTPHPDRLTPSPDDTYDDDTNHPNGRNNSIEIVPQLPPD
                     RPIIELGVATLRKNFMEASCTVETNSGLAIFWKIGNASVDAFNRGTTHTRLMRNGVPV
                     YALVSTLRVPWLNVIPLTKITCAACPTNLVAGDGVDLNSCTTKSTTIPCPGQQRTHIF
                     FSAKGDRAVCITSELVSQPTITWSVGSDRLRNDGFSQTWYGIQPGVCGILRSEVRIHR
                     TTWRFGSTSKDYLCEVSASDSKTSDYKVLPNAHSTSNFALVAATTLTVTILCLLCCLY
                     CMLTRPRASVY"
     repeat_region   667. .861
     repeat_region   892. .1101
     polyA_signal    2802. .2807
BASE COUNT      857 a   1068 c    643 g    605 t
ORIGIN
```

Query Match 5.3%; Score 149.4; DB 17; Length 3173;
Best Local Similarity 53.0%; Pred. No. 7.2e-16;
Matches 349; Conservative 0; Mismatches 301; Indels 9; Gaps 1;

```
Qy   1480 CACATCGACTGCTGCCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGC 1539
Db    459 CGCATCTACAACGACCCCAACAACAACGACCGCGGCTCCCACAACGGCCGCTACAACCAC 518

Qy   1540 TGCCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGC 1599
Db    519 AGCTGTTACTACAGCCGCGTCTACATCAGCGGAAACCACCACAGCTACTGCGACTGCTAC 578

Qy   1600 CCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGC 1659
Db    579 CTCAACCCCAACCACAACTACGCCTACGTCCACAACAACTACTACAGCTACCACCACTGT 638

Qy   1660 TCCTACTACTGCTGCCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCC 1719
Db    639 TCCAACAACCGCTTCTACAACAACTGATACGACCACAGCAGCAACGACCACAGCAGCAAC 698

Qy   1720 CACCGCAGCCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTAC 1779
Db    699 GACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAAC 758

Qy   1780 CGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCC 1839
Db    759 GACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAAC 818

Qy   1840 AGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGC 1899
Db    819 GACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAACGACCACCGCGGCTAC 878

Qy   1900 CCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCC 1959
Db    879 TACTTCCTCTGCAACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCAC 938

Qy   1960 CAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAA 2019
Db    939 CACCGCG---------GCTACCACCACCGCGGCTACCACCACCGCGGCTACCACCACCGC 989
```

```
Qy   2020 CACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACC 2079
Db    990 GGCTACCACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCACCACCGC 1049

Qy   2080 TGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2138
Db   1050 GGCTACCACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCACAACGG 1108


RESULT 15
BSU43200
LOCUS       BSU43200     2070 bp    DNA             VRT       13-MAY-1997
DEFINITION  Boreogadus saida antifreeze glycopeptide AFGP polyprotein precursor
            gene, complete cds.
ACCESSION   U43200
NID         g2078482
KEYWORDS    .
SOURCE      Boreogadus saida.
  ORGANISM  Boreogadus saida
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Paracanthopterygii; Gadiformes; Gadoidei; Gadidae; Boreogadus.
REFERENCE   1  (bases 1 to 2070)
  AUTHORS   Chen,L., DeVries,A.L. and Cheng,C.H.
  TITLE     Convergent evolution of antifreeze glycoproteins in Antarctic
            notothenioid fish and Arctic cod
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 94 (8), 3817-3822 (1997)
  MEDLINE   97268653
REFERENCE   2  (bases 1 to 2070)
  AUTHORS   Cheng,C.-H.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1995) Chi-Hing C. Cheng, Molecular and
            Integrative Physiology, University of Illinois, 524 Burrill Hall,
            407 S. Goodwin, Urbana, IL 61801, USA
FEATURES             Location/Qualifiers
     source          1. .2070
                     /organism="Boreogadus saida"
                     /db_xref="taxon:44932"
                     /clone="BS3-1"
     intron          83. .164
     sig_peptide     join(209. .211,281. .376)
                     /note="putative"
     CDS             join(209. .211,281. .1801)
                     /note="cleavage of polyprotein at conserved spacers R or
                     RAAR to produce multiple AFGP molecules"
                     /codon_start=1
                     /product="antifreeze glycopeptide AFGP polyprotein
                     precursor"
                     /db_xref="PID:g2078483"
                     /translation="MLTGSCLLGPSCTSSPRYEQHQLLVARPAAAARAATPATAATPA
                     TAATPATAATAATEATAATAATPATAATPATAATAATTAATAATAATAATPARAARAA
                     TPATAATPATAATAATAATAATAETPARAATPATAATPATAATPATAATAATAATSAT
                     AATAARAATPATAATPATPATAARAARAATPATAATAATAATAATAATAATAATAATP
                     ARAARAATPATAPTPATAATPATAATAATAPTAATPARAARAATPATAATLATAATPA
                     TPATPATAATDATAATAATPARAATPATPATAATPATPATAATAATAATAATAATPAR
                     AARAATPATAATPATAATAATAATAATAATPARAARAATPATAATAATAATAATA
                     ATAATPARAARAATPATPATPATPATPATAATAATAATAATAATAATAATAATAPTPA
                     RAARAATPATGATPATAPTAGTAATAATAATAATPARASTPATAATPVTAATPAAVF"
     intron          212. .280
     mat_peptide     377. .1798
                     /note="cleavage of polyprotein at conserved spacers R or
                     RAAR to produce multiple AFGP molecules"
                     /product="antifreeze glycopeptide AFGP polyprotein"
BASE COUNT     593 a    724 c    468 g    285 t
ORIGIN


  Query Match             5.2%;  Score 146.8;  DB 5;  Length 2070;
  Best Local Similarity   49.7%;  Pred. No. 2e-15;
  Matches  373;  Conservative    0;  Mismatches  377;  Indels    0;  Gaps    0;

Qy   1428 CTCCTGTGACCGCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGA 1487

Db    402 CCCCAGCAACAGCGGCAACTCCGGCTACTGCGGCCACTGCAGCAACAGAAGCCACAGCAG 461

Qy   1488 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1547
Db    462 CAACAGCGGCAACTCCAGCAACAGCGGCAACTCCGGCTACAGCGGCAACTGCGGCAACAA 521

Qy   1548 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1607
Db    522 CAGCAGCCACTGCAGCCACAGCAGCCACAGCAGCAACTCCAGCAAGGGCAGCAAGAGCAG 581

Qy   1608 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1667
Db    582 CCACTCCAGCAACAGCGGCAACTCCAGCAACAGCGGCCACAGCCGCCACAGCCGCCACAG 641

Qy   1668 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1727
Db    642 CAGCCACAGCAGAAACTCCAGCAAGAGCAGCCACTCCGGCAACGGCGGCAACTCCAGCAA 701

Qy   1728 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCC 1787
Db    702 CAGCGGCAACTCCGGCTACAGCGGCTACAGCGGCAACTGCGGCAACATCAGCCACAGCAG 761

Qy   1788 CCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCC 1847
Db    762 CAACTGCAGCAAGAGCAGCCACTCCAGCAACAGCGGCAACTCCGGCCACTCCGGCAACAG 821

Qy   1848 CAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACA 1907
Db    822 CGGCAAGAGCAGCAAGAGCAGCCACTCCAGCAACAGCGGCTACAGCGGCAACAGCAGCCA 881

Qy   1908 CCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCA 1967
Db    882 CTGCCGCCACAGCAGCCACCGCCGCCACTGCAGCCACAGCAGCAACTCCAGCAAGAGCAG 941

Qy   1968 CAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAG 2027
Db    942 CAAGAGCAGCCACTCCAGCAACAGCGCCAACTCCAGCAACAGCGGCAACTCCGGCAACAG 1001

Qy   2028 TCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCA 2087
Db   1002 CGGCTACAGCAGCCACTGCACCCACAGCAGCAACTCCAGCAAGAGCAGCAAGAGCAGCCA 1061

Qy   2088 ACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATG 2147
Db   1062 CTCCGGCAACAGCGGCAACTCTAGCAACAGCGGCAACTCCGGCCACTCCGGCAACACCGG 1121

Qy   2148 ACATCGACCCCCCTCTCCCCAACGACCCCA 2177
Db   1122 CCACTGCCGCCACTGACGCCACAGCCGCCA 1151


Search completed: July 18, 1999, 06:02:50
Job time: 3256 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 01:02:57 ; Search time 19.54 Seconds
(without alignments)
1082.403 Million cell updates/sec

Title: US-09-294-663-3
Perfect score: 4328
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 788

Scoring table: BLOSUM62

Searched: 74019 seqs, 26840295 residues

Database : SwissProt_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	636	14.7	5179	1	MUC2_HUMAN	Q02817 homo sapien
2	433	10.0	797	1	VGLX_HSVEB	P28968 equine herp
3	432	10.0	662	1	MUC1_XENLA	Q05049 xenopus lae
4	431	10.0	865	1	CPN_DROME	Q02910 drosophila
5	397	9.2	1255	1	MUC1_HUMAN	P15941 h mucin 1 p
6	393.5	9.1	1367	1	AMYH_YEAST	P08640 saccharomyc
7	353.5	8.2	1664	1	SLP1_CLOTM	Q06852 clostridium
8	351	8.1	474	1	VTP3_TTV1V	P19275 thermoprote
9	350.5	8.1	279	1	Y091_NPVOP	O10341 orgyia pseu
10	349.5	8.1	822	1	ANP_NOTCO	P24856 notothenia
11	331.5	7.7	400	1	MUA1_XENLA	P10667 xenopus lae
12	330.5	7.6	670	1	VG50_HSVI1	Q00130 ictalurid h
13	309	7.1	907	1	VGP3_EBV	P03200 epstein-bar
14	309	7.1	1161	1	YJ9P_YEAST	P47179 saccharomyc
15	308	7.1	860	1	CHI2_COCIM	P54197 coccidioide
16	307.5	7.1	307	1	SGS3_DROME	P02840 drosophila
17	306	7.1	267	1	EXTN_MAIZE	P14918 zea mays (m
18	301.5	7.0	886	1	VGP3_EBVA8	Q07284 epstein-bar
19	297	6.9	620	1	EXTN_TOBAC	P13983 nicotiana t
20	294.5	6.8	3149	1	TEGU_EBV	P03186 epstein-bar
21	292	6.7	283	1	EXTN_SORVU	P24152 sorghum vul
22	286	6.6	634	1	HWP1_CANAL	P46593 candida alb
23	281	6.5	1794	1	YAV1_SCHPO	Q10172 schizosacch
24	279	6.4	2476	1	ZAN_PIG	Q28983 sus scrofa
25	271	6.3	1251	1	YQU3_CAEEL	Q09550 caenorhabdi
26	270.5	6.2	263	1	SGS3_DROYA	P13728 drosophila
27	269.5	6.2	1162	1	TCNA_TRYCR	P23253 trypanosoma
28	269.5	6.2	699	1	VGLG_HSV2H	P13290 herpes simp
29	269	6.2	338	1	SED1_YEAST	Q01589 saccharomyc
30	263	6.1	600	1	SP96_DICDI	P14328 dictyosteli
31	260	6.0	467	1	RPB1_CRIGR	P11414 cricetulus
32	260	6.0	1970	1	RPB1_HUMAN	P24928 homo sapien
33	260	6.0	1970	1	RPB1_MOUSE	P08775 mus musculu
34	259.5	6.0	3421	1	TEGU_HSVEB	P28955 equine herp
35	257.5	5.9	721	1	YY21_MYCTU	Q10690 mycobacteri
36	255.5	5.9	901	1	A180_MOUSE	Q61548 mus musculu
37	253	5.8	1083	1	T2D3_HUMAN	O00268 homo sapien
38	249	5.8	1742	1	GUNA_CALSA	P22534 caldocellum
39	248	5.7	439	1	XP2_XENLA	P17437 xenopus lae
40	247.5	5.7	915	1	A180_RAT	Q05140 rattus norv
41	244	5.6	796	1	YS8A_CAEEL	Q09625 caenorhabdi
42	243	5.6	442	1	SDC3_RAT	P33671 rattus norv
43	242	5.6	1183	1	DRPL_RAT	P54258 rattus norv
44	242	5.6	217	1	SGS3_DROSI	P13729 drosophila
45	241.5	5.6	817	1	VRP1_YEAST	P37370 saccharomyc

ALIGNMENTS

```
RESULT   1
MUC2_HUMAN
ID   MUC2_HUMAN     STANDARD;      PRT;  5179 AA.
AC   Q02817; Q14878;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
GN   MUC2 OR SMUC.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE;
RX   MEDLINE; 94132002.
RA   GUM J.R. JR., HICKS J.W., TORIBARA N.W., SIDDIKI B., KIM Y.S.;
RL   J. BIOL. CHEM. 269:2440-2446(1994).
RN   [2]
RP   SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC   TISSUE=COLON;
RX   MEDLINE; 93016075.
RA   GUM J.R. JR., HICKS J.W., TORIBARA N.W., ROTHE E.-M., LAGACE R.E.,
RA   KIM Y.S.;
RL   J. BIOL. CHEM. 267:21375-21383(1992).
RN   [3]
RP   SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX   MEDLINE; 91358717.
RA   TORIBARA N.W., GUM J.R. JR., CULHANE P.J., LAGACE R.E., HICKS J.W.,
RA   PETERSEN G.M., KIM Y.S.;
RL   J. CLIN. INVEST. 88:1005-1013(1991).
CC   -!- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC       OTHER MUCUS MEMBRANE-CONTAINING ORGANS.
CC   -!- SUBUNIT: MULTIMERIC.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC       BRONCHUS, CERVIX AND GALL BLADDER.
CC   -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC       INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC       VARIES AMONG DIFFERENT ALLELES.
CC   -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC       OF SILKWORM HEMOCYTIN.
CC   -!- SIMILARITY: CONTAINS 2 VWFC DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; L21998; G454154; -.
DR   EMBL; M74027; G188864; -.
DR   EMBL; M94131; G186396; -.
DR   EMBL; M94132; G186398; -.
DR   HSSP; P15445; 1PSH.
DR   MIM; 158370; -.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 2.
KW   GLYCOPROTEIN; REPEAT; POLYMORPHISM; SIGNAL.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   5179       MUCIN 2.
FT   DOMAIN     1401   1747       APPROXIMATE REPEATS.
FT   REPEAT     1401   1416       1.
FT   REPEAT     1417   1432       2.
FT   REPEAT     1433   1448       3.
FT   REPEAT     1449   1464       4.
FT   REPEAT     1465   1471       5.
FT   REPEAT     1472   1478       6.
FT   REPEAT     1479   1494       7A.
```

```
FT   REPEAT     1495   1517       7B.
FT   REPEAT     1518   1533       8A.
FT   REPEAT     1534   1556       8B.
FT   REPEAT     1557   1572       9A.
FT   REPEAT     1573   1596       9B.
FT   REPEAT     1597   1612       10A.
FT   REPEAT     1613   1635       10B.
FT   REPEAT     1636   1651       11A.
FT   REPEAT     1652   1675       11B.
FT   REPEAT     1676   1683       12.
FT   REPEAT     1684   1699       13.
FT   REPEAT     1700   1715       14.
FT   REPEAT     1716   1731       15.
FT   REPEAT     1732   1747       16.
FT   DOMAIN     4815   4886       VWFC 1.
FT   DOMAIN     4924   4991       VWFC 2.
FT   DOMAIN     5075   5160       CTCK.
FT   DISULFID   5075   5122       BY SIMILARITY.
FT   DISULFID   5089   5136       BY SIMILARITY.
FT   DISULFID   5098   5152       BY SIMILARITY.
FT   DISULFID   5102   5154       BY SIMILARITY.
FT   DISULFID      ?   5159       BY SIMILARITY.
FT   CARBOHYD    163    163       POTENTIAL.
FT   CARBOHYD    423    423       POTENTIAL.
FT   CARBOHYD    670    670       POTENTIAL.
FT   CARBOHYD    770    770       POTENTIAL.
FT   CARBOHYD    894    894       POTENTIAL.
FT   CARBOHYD   1139   1139       POTENTIAL.
FT   CARBOHYD   1154   1154       POTENTIAL.
FT   CARBOHYD   1215   1215       POTENTIAL.
FT   CARBOHYD   1230   1230       POTENTIAL.
FT   CARBOHYD   1246   1246       POTENTIAL.
FT   CARBOHYD   1787   1787       POTENTIAL.
FT   CARBOHYD   1820   1820       POTENTIAL.
FT   CARBOHYD   4339   4339       POTENTIAL.
FT   CARBOHYD   4351   4351       POTENTIAL.
FT   CARBOHYD   4362   4362       POTENTIAL.
FT   CARBOHYD   4373   4373       POTENTIAL.
FT   CARBOHYD   4422   4422       POTENTIAL.
FT   CARBOHYD   4438   4438       POTENTIAL.
FT   CARBOHYD   4502   4502       POTENTIAL.
FT   CARBOHYD   4616   4616       POTENTIAL.
FT   CARBOHYD   4627   4627       POTENTIAL.
FT   CARBOHYD   4752   4752       POTENTIAL.
FT   CARBOHYD   4787   4787       POTENTIAL.
FT   CARBOHYD   4881   4881       POTENTIAL.
FT   CARBOHYD   4888   4888       POTENTIAL.
FT   CARBOHYD   4955   4955       POTENTIAL.
FT   CARBOHYD   4970   4970       POTENTIAL.
FT   CARBOHYD   5019   5019       POTENTIAL.
FT   CARBOHYD   5038   5038       POTENTIAL.
FT   CARBOHYD   5069   5069       POTENTIAL.
FT   CONFLICT   1351   1351       H -> L (IN REF. 3).
FT   CONFLICT   1412   1412       T -> S (IN REF. 3).
FT   CONFLICT   1449   1449       L -> P (IN REF. 3).
FT   CONFLICT   1504   1504       M -> T (IN REF. 3).
FT   CONFLICT   4192   4192       G -> S (IN REF. 2).
SQ   SEQUENCE   5179 AA;  540295 MW;  ECD04582 CRC32;


  Query Match             14.7%;  Score 636;  DB 1;  Length 5179;
  Best Local Similarity   25.6%;  Pred. No. 1.2e-20;
  Matches  201;  Conservative   41;  Mismatches  213;  Indels  330;  Gaps   24;

Qy      29 LHEPHP-------DCPPAEQHWLLPHEYDCTKF------------YYCEYGLKFIAPRD 68
           ::| ||       |  | :     | : :|                 |:  : ||   :
Db    1309 INEDHPSSGSDDGDREPFDGVCGAPEDIECRSVKDPHLSLEQHGQKVQCDVSVGFICKNE 1368

Qy      69 CAPGT-------EFKFSAQTCVHAALAGC-TLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
             |        ::|     |    :  | | | ||  |   | ||   |||  ||||
Db    1369 DQFGNGPFGLCYDYKIRVNCC--WPMDKCITTPSPPTTTPSPPPTT---TTTLPPTTTPS 1423

Qy     121 APTTTTQAPTTTTQAPTTTQAP----TTTQAPTTTQAP---TTTTQAPTTTTQAPTTTTQ 173
```

```
                ||||     |||| |||| :|    |||  |||| :|   ||||  ||||   ||||
Db    1424 PPTT-----TTTTPPPTTTPSPPITTTTTPLPTTTPSPPISTTTTPPPTTTPSPPTTTPS 1478

Qy     174 APTTTQAP----TTTQAPTTTQAPTTITQAATTPAATT--PAATTPAATTPAATTPAATT 227
           |||| :|    |||  |||| :|  :|   | ||:||  |  |||:  |   |||  ||
Db    1479 PPTTTPSPPTTTTTTPPPTTTPSP-PMTTPITPPASTTTLPPTTTPSPPTTTTTTPPPTT 1537

Qy     228 PGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLY 287
             | ||: |: ||
Db    1538 TPSP-PTTTPITPP---------------------------------------------- 1551

Qy     288 FNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNR 347

Db    1551 ---------------------------------------------------------- 1551

Qy     348 CDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYC 407

Db    1551 ---------------------------------------------------------- 1551

Qy     408 PTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCE 467

Db    1551 ---------------------------------------------------------- 1551

Qy     468 SPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPS-----TVVPP----A 518
                              |:|  ||  |:  ||   |  ||  ||     |  ||
Db    1551 ---------------TSTTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPPTITTT 1594

Qy     519 TPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPI 578
           ||| |  | ||||   || ||  |:   ||      :| |:|||  |:  |||     |
Db    1595 TPPPTTTPSPPTTTTTTPPPTTTPSPPTTTPITPPTSTTTLPPTTTPSPPPTTTTTPPPT 1654

Qy     579 TVTSAPTAAPTAAPTAAPTAAPTTAVPEIP-TTVTSPPTAAPTTAAPAPNTTVTVPPTAA 637
           |  | ||    : |    |  | |  |  | ||  ||||   ||  |:| || : | |
Db    1655 TTPSPPTTTTPSPPITTTTTPPPTTTPSSPITTTPSPPTTTMTT--PSPTTTPSSPIT-- 1711

Qy     638 PTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNT--TAAPVTTTSAP-ATTPEDDDIDPPL 694
           ||  |:  || : |||   | :| |    | |   |  | ||||:|  |||    | ||
Db    1711 -TTTTPSSTTTPSPPPTTMTTPSPTTTPSPPTTTMTTLPPTTTSSPLTTTPLPPSITPPT 1769

Qy     695 -----PNDPINPCVEECNVLPWAHADCDKYWVCDGNNQVL-VVCSEGLQFNPTTKTCDFA 748
                |  |||  ||   |  :    :    |: :::  ||  |           :|
Db    1770 FSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDTELIGDVCGPG----------WA 1818

Qy     749 CNVGC 753
           |: |
Db    1819 ANISC 1823


RESULT   2
VGLX_HSVEB
ID   VGLX_HSVEB     STANDARD;      PRT;   797 AA.
AC   P28968;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE   GLYCOPROTEIN X PRECURSOR.
GN   71.
OS   EQUINE HERPESVIRUS TYPE 1 (STRAIN AB4P) (EHV-1).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92295566.
RA   TELFORD E.A.R., WATSON M.S., MCBRIDE K., DAVISON A.J.;
RL   VIROLOGY 189:304-316(1992).
DR   EMBL; M86664; G330862; -.
DR   PIR; H36802; VGBEX1.
KW   GLYCOPROTEIN; TRANSMEMBRANE; SIGNAL.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    797       GLYCOPROTEIN X.
FT   DOMAIN       23    465       SER/THR-RICH.
FT   TRANSMEM    766    790       POTENTIAL.
FT   CARBOHYD    590    590       POTENTIAL.
SQ   SEQUENCE   797 AA;  80342 MW;  18DE40E1 CRC32;
```

```
  Query Match           10.0%;  Score 433;  DB 1;  Length 797;
  Best Local Similarity   25.1%;  Pred. No. 1.4e-12;
  Matches  181;  Conservative   40;  Mismatches  167;  Indels  334;  Gaps   21;

Qy      78 SAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTT--TTQA 135
           | |:   || :  : |   : ||  | :|   |||  || :|  ||||| ||||  || |
Db      77 STQSSSTAATSS-SAPSTASSTTSIPTSTSTETTTTTPTASTTTPTTTTAAPTTAATTTA 135

Qy     136 PTT---TQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTT--------------------TT 172
            ||   | | |||   | |  |||||   |||| | ||                     ||
Db     136 VTTAASTSAETTTATATATSTPTTTTPTSTTTTTATTTVPTTASTTTDTTTAATTTAATT 195

Qy     173 QAPTTTQAPT---TTQAPTTTQAPTT------ITQAATTPAATTPAATTPAATTPAATTP 223
           | ||| | |   || | ||| | ||       | |||| |||| |||| |||| ||||
Db     196 TAATTTAATTTAATTTAATTTAATTTAATTSSATTAATTTAATTTAATTTAATTTAATTT 255

Qy     224 AATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCP 283
           |||| |  :|||
Db     256 AATTTG--SPTS---------------------------------------------- 266

Qy     284 EGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLP 343

Db     266 ------------------------------------------------------------ 266

Qy     344 HGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACT 403
                                                 | |   | |
Db     266 ----------------------------------------GSTSTTGASTSTP------- 279

Qy     404 GWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPAT 463
                                                                 | :|
Db     279 -----------------------------------------------------SAST 282

Qy     464 QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPA- 523
            :  :|                |||:   || |: ||  ||:| |:| ||   | : |
Db     283 ATSATP---------------TSTSTSAAATTSTPT--PTSAATSAESTTEAPTSTPTT 324

Qy     523 -TAAPVPPTTAIPTPAPT---AAPTAAPTTA-------APESPT--TVTVPPTAAPTAAP 569
            |  |   |||  :|  |   |: |:| |||       :|:|  |  | |  |::  |  |
Db     325 DTTTPSEATTATTSPESTTVSASTTSATTTAFTTESHTSPDSSTGSTSTAEPSSTFTLTP 384

Qy     570 TTAVPEI-PITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNT 628
           :|| |     | :|| | : :   :  ||    :      ||  |    : |  :
Db     385 STATPSTDQFTGSSASTESDSTDSSTVPTTGTESITESSSTTEASTNLGSSTYESTEALE 444

Qy     629 TVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAPP--------------TVAHA------- 668
           |     |:  || :|:| |     |: | |   |               | |:|
Db     445 TPDGNTTSGNTTPSPSPRT-----PSFADTQQTPDNGVSTQHTTINDHTTANAQKHAGHH 499

Qy     668 --------------PNTTAAPVTTTSAPATTPEDD-----------DIDPPLPNDPINPC 702
                         :||  |   | :|  | :||           :| | ||  |  |
Db     500 RGRAGGRRGSPQGGSHTTPHPDRLTPSPDDTYDDDTNHPNGRNNSIEIVPQLP--PDRPI 557

Qy     703 VE 704
           :|
Db     558 IE 559


RESULT   3
MUC1_XENLA
ID   MUC1_XENLA     STANDARD;      PRT;   662 AA.
AC   Q05049;
DT   01-OCT-1994 (REL. 30, CREATED)
DT   01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   INTEGUMENTARY MUCIN C.1 (FIM-C.1) (FRAGMENT).
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKIN;
RX   MEDLINE; 93077556.
RA   HAUSER F., HOFFMANN W.;
RL   J. BIOL. CHEM. 267:24620-24624(1992).
CC   -!- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC       INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: SKIN.
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC   -!- ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN
CC       MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS 6 P-TYPE (TREFOIL) DOMAINS.
DR   EMBL; L02115; G951460; -.
DR   PIR; A45155; A45155.
DR   HSSP; P04002; 1ATF.
DR   PROSITE; PS00025; P_TREFOIL; 6.
KW   REPEAT; AMPHIBIAN SKIN; GLYCOPROTEIN; ALTERNATIVE SPLICING.
FT   NON_TER       1      1
FT   DOMAIN       81    144       8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT                                 ALA/THR-RICH.
FT   REPEAT       81     88       1-1.
FT   REPEAT       89     96       1-2.
FT   REPEAT       97    104       1-3.
FT   REPEAT      105    112       1-4.
FT   REPEAT      113    120       1-5.
FT   REPEAT      121    128       1-6.
FT   REPEAT      129    136       1-7.
FT   REPEAT      137    144       1-8.
FT   DOMAIN      161    202       P-TYPE 1.
FT   DOMAIN      218    301       8 X APPROXIMATE TANDEM REPEATS, THR-RICH.
FT   REPEAT      218    224       2-1.
FT   REPEAT      225    239       2-2.
FT   REPEAT      240    249       2-3.
FT   REPEAT      250    259       2-4.
FT   REPEAT      260    275       2-5.
FT   REPEAT      276    287       2-6.
FT   REPEAT      288    294       2-7.
FT   REPEAT      295    301       2-8.
FT   DOMAIN      306    347       P-TYPE 2.
FT   DOMAIN      353    394       P-TYPE 3.
FT   DOMAIN      402    522       12 X APPROXIMATE TANDEM REPEATS,
FT                                 THR-RICH.
FT   REPEAT      402    411       3-1.
FT   REPEAT      412    419       3-2.
FT   REPEAT      420    431       3-3.
FT   REPEAT      432    443       3-4.
FT   REPEAT      444    453       3-5.
FT   REPEAT      454    460       3-6.
FT   REPEAT      461    472       3-7.
FT   REPEAT      473    479       3-8.
FT   REPEAT      480    491       3-9.
FT   REPEAT      492    498       3-10.
FT   REPEAT      499    515       3-11.
FT   REPEAT      516    522       3-12.
FT   DOMAIN      525    566       P-TYPE 4.
FT   DOMAIN      572    613       P-TYPE 5.
FT   DOMAIN      620    661       P-TYPE 6.
FT   DISULFID    162    188       BY SIMILARITY.
FT   DISULFID    172    187       BY SIMILARITY.
FT   DISULFID    182    199       BY SIMILARITY.
FT   DISULFID    307    333       BY SIMILARITY.
FT   DISULFID    317    332       BY SIMILARITY.
FT   DISULFID    327    344       BY SIMILARITY.
FT   DISULFID    354    380       BY SIMILARITY.
FT   DISULFID    364    379       BY SIMILARITY.
FT   DISULFID    374    391       BY SIMILARITY.
FT   DISULFID    526    552       BY SIMILARITY.
FT   DISULFID    536    551       BY SIMILARITY.
FT   DISULFID    546    563       BY SIMILARITY.
FT   DISULFID    573    599       BY SIMILARITY.
FT   DISULFID    583    598       BY SIMILARITY.
FT   DISULFID    593    610       BY SIMILARITY.
FT   DISULFID    621    647       BY SIMILARITY.
FT   DISULFID    631    646       BY SIMILARITY.
FT   DISULFID    641    658       BY SIMILARITY.
```

```
FT   VARIANT     276    276       K -> E.
FT   VARIANT     354    354       C -> R.
FT   VARIANT     415    415       T -> A.
SQ   SEQUENCE   662 AA;  67774 MW;  00C98798 CRC32;


  Query Match             10.0%;  Score 432;  DB 1;  Length 662;
  Best Local Similarity   22.6%;  Pred. No. 1.3e-12;
  Matches  191;  Conservative   27;  Mismatches  229;  Indels  400;  Gaps   27;

Qy      65 APRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTT----QAPTT--- 118
           ||   | |     :| |    | ||     |   | :||||  ||  |    :||||
Db      73 APATTAAGKAPTTAAATAPTTAAAG----APTTATGKAPATAAAPVPTTAASKAPTTAAA 128

Qy     118 ------TTQAPTTTTQA---------------------------------------------- 129
                 ||||   |
Db     129 ATHSTAAAAAPTTAASAAKSKERSTSSSSEEEHCHVKPSKREMCGSKGITKKQCKKKNCC 188

Qy     129 ------------------------PTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTT 164
                                    ||||:|||| |  |||  ||||   |  |   |||
Db     189 FDPKGHGGIHCFHRKPKGHSHEEHTTTTTKAPTTIQIATTTTTPTTTTTTTKATPTTTTT 248

Qy     165 TQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPA 224
           |:|  |||   |||:| |||  ||||   || |:| |||  ||   |||  ||  |||
Db     249 TKATPTTT---TTTKATTTTTTPTTT---TTTTKATTTPTTTT--TTTPTTTTTKATTTT 300

Qy     225 ATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPE 284
             || |
Db     301 TTTSG---------------------------------------------------------- 306

Qy     285 GLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPH 344
                              ||  | |
Db     306 ----------------ECKMEPS---------------------------------------- 313

Qy     345 GNRCDKYYQCVHGNLVERRCGA-GTHFSFELQQ---CDHIELVGCTLPGGESEEVDVDED 400
                |   | :  : | :|   |  |   : |   | :      ||    |:  |
Db     313 -----KREDCGYSGITESQCRTKGCCFDSSIPQTKWCFY------TL----SQVAD---- 354

Qy     401 ACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFS 460
               | || :                 |||
Db     354 ------CKVEPSQ--------------------RVDCG---------------------- 366

Qy     461 PATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATP 520
                  |    :|   | |  ::      :      : :  ||  |   | |
Db     366 ----------FRGITADQCRQKNCCFDSSI-----SGTKWCFYSTSQVAATKTTTTPTT- 410

Qy     521 PATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITV 580
                 |||   | | |  ||   ||     ||| |
Db     410 -----TTTPTTTTTTKATTTTPTTTTTT-----PTTTTT---------------------- 439

Qy     581 TSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTT 640
                  |   | | |  |||  |   || |  |  |||    | || |   |  |||
Db     439 --------TTTTKATTTTPTTTTP--TTTTTKATTTTPTTTTTTPTTTTTKATTTTPTT 488

Qy     641 AAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPIN 700
             | || |   |  ||     |      || |  ||||          |   |    :
Db     489 TTTTPTTTTTKATTTTPTTTTTT----TTTKATTTTSGECKMEPSKRADCGYPGITES 544

Qy     701 PC----------------------VEECNVLPWAHADC---------------------- 717
            |                      | :| | | :  ||
Db     545 QCRSKGCCFDSSIPQTKWCFYSLPQVADCKVAPSSRVDCGFGGITADQCRQRNCCFDSSI 604

Qy     717 -DKYWVCDGNNQVLVVCSEGLQFNPTTKTCD---------FACNVGCVRSNIQMSESYEG 766
             |     :|   :||     | ||  |           | ||   |  |:
Db     605 SGTKWCFYSTSQGNAMCS-----GPPTKRRDCGYPGISSSVCINRGCCWDNSVMN----- 655

Qy     767 VQVFIPW 773
               :||
Db     655 ----VPW 657


RESULT   4
CPN_DROME
```

```
ID   CPN_DROME      STANDARD;      PRT;   865 AA.
AC   Q02910;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE   CALPHOTIN.
GN   CPN OR CAP.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE; 93165729.
RA   MARTIN J.H., BENZER S., RUDNICKA M., MILLER C.A.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 90:1531-1535(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE; 93165730.
RA   BALLINGER D.G., XUE N., HARSHMAN K.D.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 90:1536-1540(1993).
CC   -!- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
CC       REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
CC       OF CA+2 PER MOL OF PROTEIN.
CC   -!- SUBUNIT: HOMODIMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC   -!- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
CC       COMPOUND EYES AND OCELLI.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
CC       DEVELOPMENT.
DR   EMBL; L02111; G157032; -.
DR   EMBL; L05080; G157072; -.
DR   PIR; A47282; A47282.
DR   FLYBASE; FBgn0010218; Cpn.
KW   CALCIUM-BINDING.
FT   CONFLICT     36     36       A -> AVAPAVVA (IN REF. 2).
FT   CONFLICT     43     43       I -> T (IN REF. 2).
FT   CONFLICT     64     64       I -> V (IN REF. 2).
FT   CONFLICT     76     76       T -> A (IN REF. 2).
FT   CONFLICT    100    100       P -> PP (IN REF. 2).
FT   CONFLICT    126    127       VQ -> AP (IN REF. 2).
FT   CONFLICT    154    154       I -> V (IN REF. 2).
FT   CONFLICT    160    160       S -> T (IN REF. 2).
FT   CONFLICT    534    534       A -> E (IN REF. 2).
FT   CONFLICT    699    699       I -> T (IN REF. 2).
FT   CONFLICT    703    703       V -> L (IN REF. 2).
FT   CONFLICT    721    721       D -> E (IN REF. 2).
SQ   SEQUENCE   865 AA;  84781 MW;  7C1F27FB CRC32;


  Query Match             10.0%;  Score 431;  DB 1;  Length 865;
  Best Local Similarity   25.8%;  Pred. No. 1.8e-12;
  Matches  183;  Conservative   43;  Mismatches  248;  Indels  234;  Gaps   32;

Qy      91 TLPGPPAETTQAPAT---TQAPTTTQAPTTTQAPT-------------TTTQAPTTTT 133
           |:| | :    || |     ||    :|     ||                 : |  | |
Db       5 TIPSPVSAPVAAPVTPSAVAAPVQVVSPAAVAPAPAAPIAVTPVAPPPTLASVQPATVTI 64

Qy     134 QAPTTTQA----PTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTT 189
             ||    |    |· : ||    |||    :| :|   |    | |    ||    ||
Db      65 PAPAPIAAASVTPVASVAPPVVAAPTPPAASPVST---PVAVAQIPVAVSAPV---APPV 118

Qy     190 TQAPTTITQ-AATTPAATTP----AATTPAATTPAATTPAATTPGVPAPTSAPVWPPICE 244
             || : |    |   ||    :| |||| ||  :   |: | ||| |: ||  |:
Db     119 AATPTPVVQIPVAAPVIATPPVAASAPTPAAVTPVISPVIASPPVVPANTTVPVAAPVA- 178

Qy     245 LLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECD 304
             :|   |                                        | :    :||
Db     178 AVPAAVPV----------------------------------VAPVLAPAVAPAVAPVV 202

Qy     305 GEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRC 364
           |    ||| |      | :  : :   |       |:|  :
Db     203 AETPAPPPVAE------IPVATIPECVAP-------LIPEVS----------------- 232
```

```
Qy    365 GAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPAD 424
                          |  | |   :| | |   |       ||       |   ||
Db    232 -----------------VVATKPLAAAEPVVVAPPA-------TE------TPVVAPAA 260

Qy    425 FSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQ 484
           |     || |                 |       :| : | | || |
Db    261 AS-----PHVS--------------VAPAVETAVVAPVSASTEPPVAA------------ 290

Qy    485 CTSTAAPTAAPTAAPTAA-------------PTAAPTAAPSTVVPP-------------- 518
           | | ||   |  ||  |              || ||   | |: ||
Db    290 ATLTTAP-ETPALAPVVAESQVAANTVVATPPTPAP--EPETIAPPVVAETPEVASVAVA 346

Qy    518 -ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVT-----VPPTAAPTAAPTT 571
            |||      |||  |   |||  | |  | | |  |         :||  ||:  | :
Db    347 ETTPPV----VPPVAAESIPAPVVATTPVPATLAVTDPDVTASAVPELPPVIAPSPVP-S 401

Qy    572 AVPEIPITVTSAPTAAPTAA------------PTAAPTAAPTTA----VPEIPTTVTSP- 615
           || || |: : : |   | ||              | || :|| |    :||:   : :|
Db    402 AVAETPVDL-APPVLPPVAAEPVPAVVAEETPETPAPASAPVTIAALDIPEVAPVIAAPS 460

Qy    615 --PTAAPTTAAP---APNTTVTVPPTAAPTTAAPAPNTTVTVPPTAA---PTAAPPTVAH 666
            | :||: |||    | || :|| | ||  | |     |:|   ||   | | |  |
Db    461 DAPAEAPSAAAPIVSTPPTTASVPETTAPPAAVPTEPIDVSVLSEAAIETPVAPPVEV-- 519

Qy    667 APNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHA 714
             ||   |    ||     |  |:|  |  ||   :|:   :|   |
Db    519 ---TTEVAVADV-APPEAAADLIIEPVEPPAPIPDLLEQTTSVPAVEA 562


RESULT   5
MUC1_HUMAN
ID   MUC1_HUMAN     STANDARD;      PRT;  1255 AA.
AC   P15941; P15942; P13931; P17626;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)
DE   (EPISIALIN) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
DE   (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG)
DE   (PEANUT-REACTIVE URINARY MUCIN) (PUM).
GN   MUC1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PANCREAS;
RX   MEDLINE; 90368716.
RA   LAN M.S., BATRA S.K., QI W.-N., METZGAR R.S., HOLLINGSWORTH M.A.;
RL   J. BIOL. CHEM. 265:15294-15299(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90202794.
RA   LIGTENBERG M.J.L., VOS H.L., GENNISSEN A.M.C., HILKENS J.;
RL   J. BIOL. CHEM. 265:5573-5578(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST CARCINOMA;
RX   MEDLINE; 90368715.
RA   GENDLER S.J., LANCASTER C.A., TAYLOR-PAPADIMITRIOU J., DUHIG T.,
RA   PEAT N., BURCHELL J., PEMBERTON L., LALANI E.-N., WILSON D.;
RL   J. BIOL. CHEM. 265:15286-15293(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91097524.
RA   LANCASTER C.A., PEAT N., DUHIG T., WILSON D.,
RA   TAYLOR-PAPADIMITRIOU J., GENDLER S.J.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 173:1019-1029(1990).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST CARCINOMA;
RX   MEDLINE; 90276413.
RA   WRESCHENER D.H., HAREUVENI M., TSARFATY I., SMORODINSKY N.,
RA   HOROV J., ZARETSKY J., KOTKES P., WEISS M., LATHE R., DION A.,
RA   KEYDAR I.;
RL   EUR. J. BIOCHEM. 189:463-473(1990).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST CARCINOMA;
RX   MEDLINE; 90276414.
RA   HAREUVENI M., TSARFATY I., ZARETSKY J., KOTKES P., HOREV J.,
RA   ZRIHAN S., WEISS M., GREEN S., LATHE R., KEYDAR I., WRESCHNER D.H.;
RL   EUR. J. BIOCHEM. 189:475-486(1990).
RN   [7]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91033045.
RA   TSARFATY I., HAREUVENI M., HOREV J., ZARETSKY J., WEISS M.,
RA   JELTSCH J.M., GARNIER J.M., LATHE R., KEYDAR I., WRESCHNER D.H.;
RL   GENE 93:313-318(1990).
RN   [8]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE; 88330762.
RA   GENDLER S.J., TAYLOR-PAPADIMITRIOU J., DUHIG T., ROTHBARD J.,
RA   BURCHELL J.;
RL   J. BIOL. CHEM. 263:12820-12823(1988).
RN   [9]
RP   SEQUENCE OF 1-46 FROM N.A.
RC   TISSUE=BREAST CARCINOMA;
RA   BULUWELA L., LIU Q., LUQMANI Y.A., GOMM J.J., COOMBES R.C.;
RL   SUBMITTED (OCT-1992) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC       CYTOSKELETON.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
CC       IS ALSO PRODUCED.
CC   -!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL
CC       TUMORS, SUCH AS BREAST CANCER.
CC   -!- PTM: HIGHLY GLYCOSYLATED (N- & O-LINKED CARBOHYDRATES & SIALIC
CC       ACID).
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC       VARIES AMONG DIFFERENT ALLELES. THE BREAST TUMOR MUC1 HAS
CC       GENERALLY ONLY 5 TO 7 REPEATS.
CC   -!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE
CC       SPLICING OF THE MUC1 GENE.
DR   EMBL; J05582; G189599; -.
DR   EMBL; M32738; G182124; -.
DR   EMBL; M32739; G182129; -.
DR   EMBL; J05581; G188870; -.
DR   EMBL; M61170; G189784; -.
DR   EMBL; X52229; G37054; ALT_SEQ.
DR   EMBL; X52228; G36435; ALT_SEQ.
DR   EMBL; M35093; G182253; ALT_SEQ.
DR   EMBL; Z17324; G34807; -.
DR   EMBL; Z17325; G34809; -.
DR   EMBL; M21868; G386963; ALT_SEQ.
DR   PIR; A35175; A35175.
DR   PIR; B35175; B35175.
DR   PIR; S10218; S10218.
DR   MIM; 158340; -.
KW   GLYCOPROTEIN; SIGNAL; CYTOSKELETON; ACTIN-BINDING; TRANSMEMBRANE;
KW   REPEAT; ALTERNATIVE SPLICING; POLYMORPHISM.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24   1255       MUCIN 1.
FT   DOMAIN       24   1162       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1163   1186       POTENTIAL.
FT   DOMAIN     1187   1255       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       81    960       44 X 20 AA TANDEM REPEATS.
FT   CARBOHYD    957    957       POTENTIAL.
FT   CARBOHYD    975    975       POTENTIAL.
FT   CARBOHYD   1029   1029       POTENTIAL.
FT   CARBOHYD   1055   1055       POTENTIAL.
FT   CARBOHYD   1133   1133       POTENTIAL.
FT   VARSPLIC     19     19       T -> ATTAPKPAT (IN FORM B).
FT   VARSPLIC     20     22       MISSING (IN FORM C).
FT   VARSPLIC     20     31       MISSING (IN FORM D).
FT   VARSPLIC    126    905       MISSING (IN POLYMORPHIC EPITHELIAL
```

```
FT                                VARIANT).
FT   VARSPLIC   1077   1087       FLQIYKQGGFL -> VSIGLSFPMLP (IN SECRETED
FT                                FORM).
FT   VARSPLIC   1088   1255       MISSING (IN SECRETED FORM).
FT   CONFLICT   1021   1021       S -> T (IN REF. 3).
FT   CONFLICT   1251   1251       A -> T (IN REF. 3).
SQ   SEQUENCE   1255 AA;  122072 MW;  B06B1004 CRC32;


  Query Match             9.2%;  Score 397;  DB 1;  Length 1255;
  Best Local Similarity   26.1%;  Pred. No. 6.9e-11;
  Matches  199;  Conservative   42;  Mismatches  288;  Indels  234;  Gaps   34;

Qy      10 ALGLVA--ARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYGLKFIAPR 67
           |||      |    |  :  ||   |    |||                     :|:
Db     107 ALGSTTPPAHDVTSAPDNKPA---PGSTAPPA------------------HGVTSAPDT 144

Qy      68 DCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTT-------TQ 120
             |||:    :|       |  | | | :    |    | || |  || :|       |
Db     145 RPAPGS----TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 200

Qy     121 APTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTTQAPTTTTQAPTTTQ 179
           || |     :|   |    | || |  || :|  |    | || |     :|   |    |
Db     201 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 260

Qy     180 APTTTQAPTTTQAPT-TITQAATT---PAATTPAA---TTPAATTPAATTPAATTPGVPA 232
           || |  || :|  |   :| | |   | :| | |   |:   | ||   | :| |
Db     261 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA---PGSTAPPAHG 317

Qy     233 PTSAP--------VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTF--EQRC 282
           ||||          ||                          ::| |   : |
Db     318 VTSAPDTRPAPGSTAPP-----------------------------AHGVTSAPDTRP 346

Qy     283 PEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLL 342
             |    |      :|      |  | |||
Db     347 APGSTAPPAHGVTSAPDTRPAPG--STAPP------------------------------ 375

Qy     343 PHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDAC 402
                   ||                       |     | | |          |
Db     375 ------------AHG---------------VTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 407

Qy     403 TGWYCP-----TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCP 454
             |   |     |   :  | |                    |     || | |   || |
Db     408 PGSTAPPAHGVTSAPDTRPAP------------------GSTAPPAHGVTSAPDTRPAP 448

Qy     455 GNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTV 514
           |:    ||     :|          |:     ||| |     |:||    |    || |:
Db     449 GST-APPAHGVTSAP----------DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 497

Qy     515 V--PPATPPATAAPVPP---TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAP 569
           |   | | ||  :  ||    |: |   |    ||'|      :| |   | : || |
Db     498 VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 557

Qy     570 TTAVPEI-PITVTSAP-----TAAPTAAPTAAPTAAPTTAVPEIPTT-----VTSPP--- 616
           |: |:  |   ::||      |:||    |     || |    |    | |       |:||
Db     558 VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 617

Qy     616 -TAAPTTAAPAPNTTVTVPP----TAAPTTAAPAPNTTVTVPPTAAPTAAPPT------- 664
           |:|| |  ||| :  | ||    |:|| |  ||| :  | ||    |:|| |
Db     618 VTSAPDT-RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPAHGVTSAPDTRPAPGST 671

Qy     664 ------VAHAPNTTAAPVTT-------TSAPATTPEDDDIDPP 693
                 |  ||:|  || :|       |||| | |      ||
Db     672 APPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 714


RESULT   6
AMYH_YEAST
ID   AMYH_YEAST     STANDARD;      PRT;  1367 AA.
AC   P08640; P08068;
DT   01-AUG-1988 (REL. 08, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   GLUCOAMYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-
DE   GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).
GN   STA1 OR STA2 OR MAL5 OR YIR019C.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA   CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
RA   GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA   LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA   RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA   WALSH S.V., WHITEHEAD S.;
RL   SUBMITTED (DEC-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX   MEDLINE; 87194600.
RA   YAMASHITA I., NAKAMURA M., FUKUI S.;
RL   J. BACTERIOL. 169:2142-2149(1987).
RN   [3]
RP   SEQUENCE OF 1-31 FROM N.A.
RC   STRAIN=SPX101-1C;
RX   MEDLINE; 89031230.
RA   PARDO J.M., IANEZ E., ZALACAIN M., CLAROS M.G., JIMENEZ A.;
RL   FEBS LETT. 239:179-184(1988).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
CC       GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
CC       WITH RELEASE OF BETA-D-GLUCOSE.
DR   EMBL; Z47047; G763364; -.
DR   EMBL; Z38061; G557822; -.
DR   EMBL; M16164; G172525; -.
DR   EMBL; M16165; G172526; -.
DR   EMBL; X13857; G4552; -.
DR   PIR; B26877; B26877.
DR   PIR; A26877; A26877.
DR   PIR; S48478; S48478.
DR   SGD; L0002104; STA1.
KW   HYDROLASE; GLYCOSIDASE; POLYSACCHARIDE DEGRADATION; GLYCOPROTEIN;
KW   SIGNAL; MULTIGENE FAMILY.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1367       GLUCOAMYLASE S1/S2.
FT   DOMAIN      210   1367       SER/THR-RICH.
FT   CARBOHYD    817    817       POTENTIAL.
FT   CARBOHYD    874    874       POTENTIAL.
SQ   SEQUENCE   1367 AA;  136110 MW;  212ED195 CRC32;


  Query Match             9.1%;  Score 393.5;  DB 1;  Length 1367;
  Best Local Similarity   19.8%;  Pred. No. 1e-10;
  Matches  150;  Conservative  120;  Mismatches  229;  Indels  259;  Gaps   28;

Qy      32 PHPDCPPAEQHWLLPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCT 91
           || |  |             |||        |    : |   |       :      :
Db     288 PHHDTTP-----------CTK-------KKTTTSKTCTKKTTTPVPTPSSSTTESSSAP 328

Qy      92 LPGPPAETTQ---APATTQAPTTTQAPTTTQAPTT------TTQAPTTTTQAPTTTQAP 142
           :| | : ||:   || |:      :: ||  |  : ||       | : | :: || |:
Db     329 VPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTT 388

Qy     143 TTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTT------TQAPTTTQAPTTTQAPTT- 196
           :: ||  | : :||  :    | : | :: || |       : || |:    :: || |
Db     389 ESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTS 448

Qy     196 -ITQAATTPAATTPAATTPAATTP--AATTPAATTPGVPAP-------TSAPVWPPICEL 245
            |:::: |  |  ::|| ::: |  ::|| ::: | || |        :||||
Db     449 STTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAP-VPTPSSSTTESSSAPV------- 501

Qy     246 LPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDG 305
                                        ::  |    |          :|::   :
Db     501 -------------------------TSSTTESSSAP-----------VPTPSSSTTES 522

Qy     306 EISPAP-PVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRC 364
```

```
Db    523 SSAPAPTPSSSTTESSSAPV---------------------------------------- 543

Qy    365 GAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPAD 424
Db    543 ----------------------TSSTTESSSAPV-----------------PTPSSSTTE 563

Qy    425 FSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQ 484
Db    564 SS----------------------STPVT-----SSTTESSSAPVPT-------PSSST 588

Qy    485 CTSTAAPTAAP-------TAAPTAAPTAAPTAAPSTVVPPATPPATAAPVP-------PT 530
Db    589 TESSSAPVPTPSSSTTESSSAPAPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTES 648

Qy    531 TAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPIT-----VTSAPT 585
Db    649 SSAPVPTPSSSTTESSSAPVP-TPSSSTTESSSAPVTSSTTESSSAPVPTSSSTTESSAPV 707

Qy    586 AAPTAAPT---AAPTAAPTTAVPE-------IPTTVTSPPTAAPTTAAPAPNTTVTVPPT 635
Db    708 PTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVPTP 767

Qy    636 AAPTT---AAPAPN-TTVTVPPTAAPTAAP------PTVAHAPNTTAAPVTTTSAPATTP 685
Db    768 SSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSVAPVPTPSSSSNITSSAPSSTP 827

Qy    686 -----EDDDIDPPLPND--------PINPCVEECNVLP 710
Db    828 FSSSTESSSVPVPTPSSSTTESSSAPVSSSTTESSVAP 865


RESULT   7
SLP1_CLOTM
ID   SLP1_CLOTM     STANDARD;      PRT;  1664 AA.
AC   Q06852;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER
DE   PROTEIN 1).
GN   OLPB.
OS   CLOSTRIDIUM THERMOCELLUM.
OC   PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCIB 10682;
RX   MEDLINE; 93209931.
RA   FUJINO T., BEGUIN P., AUBERT J.-P.;
RL   J. BACTERIOL. 175:1891-1899(1993).
CC   -!- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC   -!- SUBCELLULAR LOCATION: CELL WALL.
CC   -!- SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.
DR   EMBL; X67506; G296881; -.
DR   PROSITE; PS01072; SLH_DOMAIN; 2.
KW   CELL WALL; S-LAYER; SIGNAL; REPEAT.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29   1664       CELL SURFACE GLYCOPROTEIN 1.
FT   DOMAIN       36    763       4 X 156 AA APPROXIMATE REPEATS.
FT   REPEAT       36    191       1.
FT   REPEAT      207    363       2.
FT   REPEAT      409    565       3.
FT   REPEAT      607    763       4.
FT   DOMAIN      771   1377       APPROXIMATE TANDEM REPEATS OF
FT                                T-P-S-D-E-P.
FT   DOMAIN     1378   1449       GLY/PRO/SER/THR-RICH.
FT   DOMAIN     1453   1494       SLH 1 (INCOMPLETE).
FT   DOMAIN     1495   1565       SLH 2.
FT   DOMAIN     1566   1625       SLH 3.
FT   DOMAIN     1626   1646       SLH 4 (INCOMPLETE).
SQ   SEQUENCE   1664 AA;  178194 MW;  BFEC7D6A CRC32;
```

```
  Query Match             8.2%;  Score 353.5;  DB 1;  Length 1664;
  Best Local Similarity   23.5%;  Pred. No. 6.2e-09;
  Matches  149;  Conservative   70;  Mismatches  277;  Indels  139;  Gaps   23;

Qy     93 PGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQ 152
Db    917 PTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSD 976

Qy    153 APTTTTQ-------APTTTTQAP---TTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATT 202
Db    977 EPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1036

Qy    203 PAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPH 262
Db   1037 DEPTPSDEPTP-SETPEEPIPTDTPSDEPTPSDEPT--PSDEPTPSDEPTPSD------- 1087

Qy    263 DKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDI 322
Db   1087 --------EPTPSETPEEPIP----------TDTPSD-----EPTPSDEPTPSDEPTPS 1122

Qy    323 DIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIEL 382
Db   1123 DEPTPSDEPTPSETP-EEPIPTDTPSD-----------------------------EP 1150

Qy    383 VGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQ 442
Db   1151 TPSDEPTPSDEPTPSDEPTPSDEPTPSETPE-EPIPTDTPSDEPTPSDEPTPSD------ 1204

Qy    443 CVHGQTIARPCPGNLHFSPA-----TQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTA 497
Db   1204 --------EPTPSD-EPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPT--PSD 1252

Qy    498 APTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTV 557
Db   1253 EPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT-- 1311

Qy    558 TVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSP-PT 616
Db   1311 ---PSDEPTPSET---PEEPI-----PTDTPSDEPT--PSDEPTPS--DEPTPSDEPTPS 1355

Qy    617 AAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAP---------------P 662
Db   1356 DEPT---PSDEPTPSDEPTPSETPEEPTPTTTPTPTPSTTPTSGSGGSGGSGGGGGGGGG 1412

Qy    663 TVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPND 697
Db   1413 TVPTSPTP-----TPTSKPTSTPAPTEIEEPTPSD 1442


RESULT   8
VTP3_TTV1V
ID   VTP3_TTV1V     STANDARD;      PRT;   474 AA.
AC   P19275;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE   VIRAL PROTEIN TPX.
OS   THERMOPROTEUS TENAX VIRUS 1 (STRAIN VT3) (TTV1).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; LIPOTHRIXVIRIDAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90245666.
RA   NEUMANN H., ZILLIG W.;
RL   NUCLEIC ACIDS RES. 18:2171-2171(1990).
DR   EMBL; X14717; E1192689; -.
DR   PIR; S06686; S06686.
DR   PIR; S15921; S15921.
KW   REPEAT.
FT   DOMAIN      278    467       3 THR-PRO REPEATS REGIONS AND TWO NEAR
FT                                IDENTICAL REPEATS.
FT   REPEAT      278    367       THR-PRO(N).
FT   REPEAT      368    377
FT   REPEAT      378    436       THR-PRO(N).
```

```
FT   REPEAT      437    446
FT   REPEAT      447    467       THR-PRO(N).
SQ   SEQUENCE   474 AA;  49507 MW;  BAC5D831 CRC32;


  Query Match             8.1%;  Score 351;  DB 1;  Length 474;
  Best Local Similarity   32.6%;  Pred. No. 3.1e-09;
  Matches  103;  Conservative   20;  Mismatches  129;  Indels   64;  Gaps   13;

Qy     394 EVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPC 453
           |:|: : |       |:|:      ||  |:  :          |  :  | |     |
Db     192 EIDISQFAAF-----TDPLVMYLSINGSTANVKV-------YKQGSNIGTVSGNYSTTPY 239

Qy     454 PGN-----------------LHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAP 495
           ||                  | : |  |   :|:          |    | |  ||  |
Db     240 -GNPSMAGYGTVDKHYANFIVLPYEPDPQVTVTPI----------SSPSPTPTPTPTPTP 288

Qy     496 TAAPTAAPTAAPTAAPS---TVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPE 552
           |  ||  ||  ||  |:   |  |  ||  | | |  |  ||| ||  ||  ||
Db     289 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTP 348

Qy     553 SPT-TVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPT-T 610
           :|| | |  ||  ||  ||  :  :   || :||  ||  ||  ||  ||      || |
Db     349 TPTPTPTPTPTPTPTPTPTYDITYVVFDVTPSPTPTPTPTPTPTPTPT------PTPT 402

Qy     611 VTSPPTAAPTTAAPAPNTTVTVPPTAAPT-TAAPAPNTTVT-----VPPTAAPTAAPPTV 664
           |  ||  ||   | |  | |  ||  || |  | |   :|     | |:  ||   ||
Db     403 PTPTPTPTPT---PTPTPTPTPTPTPTPTPTPTPTYDITYVIFDVTPSPTPT---PTP 456

Qy     665 AHAPNTTAAPVTTTSA 680
           |  |  | :|||:
Db     457 TPTPTPTPTPTSTTSS 472


RESULT   9
Y091_NPVOP
ID   Y091_NPVOP     STANDARD;      PRT;   279 AA.
AC   O10341;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 29.3 KD PROTEIN (ORF92).
OS   ORGYIA PSEUDOTSUGATA MULTICAPSID POLYHEDROSIS VIRUS (OPMNPV).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EUBACULOVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97271300.
RA   AHRENS C.H., RUSSELL R.R., FUNK C.J., EVANS J., HARWOOD S.,
RA   ROHRMANN G.F.;
RL   VIROLOGY 229:381-399(1997).
CC   -!- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV.
DR   EMBL; U75930; G1911338; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   279 AA;  29289 MW;  AABB6954 CRC32;


  Query Match             8.1%;  Score 350.5;  DB 1;  Length 279;
  Best Local Similarity   34.2%;  Pred. No. 2.2e-09;
  Matches  102;  Conservative   34;  Mismatches  113;  Indels   49;  Gaps   14;

Qy     484 QCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPT 543
           |   |  |  :||  || :||  || :|:   |  ||  |  | |  :  ||||  :  ||
Db      26 QNIETKPPLPSPTPTPTPSPTPTPTPSPT---PTPTPTPTPTPTPTPSPTPTPALSPTPT 82

Qy     544 AAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTTA 603
           :|| :   |||  |  || :|| :||             || :|| :||  || :|| :
Db      83 PSPTLSPTPSPTP-TPSPTPSPTPSPT------------PTPSPTPSPT--PTPSPTPS 126

Qy     604 VPEIPTTVTSP-PTAAPTTAAPAPNTTVTVPPTAAPT-TAAPAPNTTVTVPPTAAPTAAP 661
               ||   || || :| | :| |: | |  || :|| | :| |: | |  || :||  |
Db     127 ----PTPTPSPTPTPSP-TPSPTPSPTPTPSPTPSPTPTPSPTPSPTPTPSPTPSPTPPP 181

Qy     662 -PTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHADCDKYW 720
```

```
           ||   :|:    |:   |:  || :  |   || |    |   |   :||   || :
Db     182 SPTPPPSPSPLGDPMYFPSSVGTTDQLRDYIRPLCNGWPRP--RETYAVPW---DCQRVM 236

Qy     721 VCDGNNQVL--VVCSEGLQFNPTTKTCDFACNVGCVRSNIQMSESYEGVQVFIPWNKL 776
           |   :  |  | |::   ::  |          |||         |:    | | | |
Db     237 HCSYFSLPLHEVTCNDDSAYSFTAD--------GCV-------PHYQSDCPFYPLNTL 279


RESULT  10
ANP_NOTCO
ID   ANP_NOTCO      STANDARD;      PRT;   822 AA.
AC   P24856;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   ANTIFREEZE GLYCOPEPTIDE POLYPROTEIN AFGP7/AFGP8 PRECURSOR.
GN   AFGP8.
OS   NOTOTHENIA CORIICEPS NEGLECTA (BLACK ROCKCOD) (YELLOWBELLY ROCKCOD).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;
OC   OSTEICHTHYES; ACTINOPTERYGII; PERCIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91067687.
RA   HSIAO K.-C., CHENG C.-H.C., FERNANDES I.E., DETRICH H.W. III,
RA   DEVRIES A.L.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 87:9265-9269(1990).
CC   -!- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED BY THE LIVER AND SECRETED INTO
CC       THE BLOOD FROM WHICH THEY BECOME DISTRIBUTED TO ALMOST THE ENTIRE
CC       EXTRACELLULAR SPACE.
CC   -!- PTM: THE DISACCHARIDE GALACTOSE-N-ACETYLGALACTOSAMINE IS ATTACHED
CC       TO THREONINES IN AFGP8 AND AFGP7.
CC   -!- CONTAINS 44 COPIES OF AFGP8 AND TWO COPIES OF AFGP7.
DR   EMBL; M55000; G213392; -.
DR   PIR; A38420; A38420.
DR   HSSP; P04002; 1ATF.
KW   ANTIFREEZE PROTEIN; GLYCOPROTEIN; POLYPROTEIN; REPEAT;
KW   MULTIGENE FAMILY; SIGNAL.
FT   SIGNAL        1      ?       POTENTIAL.
FT   PROPEP        ?     37
FT   CHAIN        38    822       ANTIFREEZE GLYCOPEPTIDE POLYPROTEIN.
FT   PEPTIDE      38     51       AFGP8.
FT   PEPTIDE      55     68       AFGP8.
FT   PEPTIDE      72     85       AFGP8.
FT   PEPTIDE      89    102       AFGP8.
FT   PEPTIDE     106    119       AFGP8.
FT   PEPTIDE     123    136       AFGP8.
FT   PEPTIDE     140    153       AFGP8.
FT   PEPTIDE     157    170       AFGP8.
FT   PEPTIDE     174    187       AFGP8.
FT   PEPTIDE     191    204       AFGP8.
FT   PEPTIDE     208    221       AFGP8.
FT   PEPTIDE     225    238       AFGP8.
FT   PEPTIDE     242    255       AFGP8.
FT   PEPTIDE     259    272       AFGP8.
FT   PEPTIDE     276    289       AFGP8.
FT   PEPTIDE     293    306       AFGP8.
FT   PEPTIDE     310    323       AFGP8.
FT   PEPTIDE     327    340       AFGP8.
FT   PEPTIDE     344    357       AFGP8.
FT   PEPTIDE     361    374       AFGP8.
FT   PEPTIDE     378    391       AFGP8.
FT   PEPTIDE     395    408       AFGP8.
FT   PEPTIDE     412    425       AFGP8.
FT   PEPTIDE     429    442       AFGP8.
FT   PEPTIDE     446    459       AFGP8.
FT   PEPTIDE     463    476       AFGP8.
FT   PEPTIDE     480    493       AFGP8.
FT   PEPTIDE     497    510       AFGP8.
FT   PEPTIDE     514    527       AFGP8.
FT   PEPTIDE     531    544       AFGP8.
FT   PEPTIDE     548    561       AFGP8.
FT   PEPTIDE     565    578       AFGP8.
```

```
FT   PEPTIDE     582    595       AFGP8.
FT   PEPTIDE     599    612       AFGP8.
FT   PEPTIDE     616    629       AFGP8.
FT   PEPTIDE     633    646       AFGP8.
FT   PEPTIDE     650    663       AFGP8.
FT   PEPTIDE     667    680       AFGP8.
FT   PEPTIDE     684    697       AFGP8.
FT   PEPTIDE     701    714       AFGP8.
FT   PEPTIDE     718    731       AFGP8.
FT   PEPTIDE     735    748       AFGP8.
FT   PEPTIDE     752    768       AFGP7.
FT   PEPTIDE     772    788       AFGP7.
FT   PEPTIDE     792    805       AFGP8.
FT   PEPTIDE     809    822       AFGP8-LIKE.
SQ   SEQUENCE   822 AA;  74643 MW;  5C3109BA CRC32;


  Query Match             8.1%;  Score 349.5;  DB 1;  Length 822;
  Best Local Similarity   26.1%;  Pred. No. 5.4e-09;
  Matches  177;  Conservative   23;  Mismatches  298;  Indels  179;  Gaps   25;

Qy      82 CVHAALAGC-----TLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAP 136
           ||  ::  |       |   | |   |||         | |  | |       | |  | ||
Db      21 CVCVSVCVCETAVTAAPAATAATAATPATAALNFAATAATPATPATPALIFAATAATAAT 80

Qy     137 TTTQA---PTTTQAPTTTQAPTTTTQAPTTTTQAPTT-------TTQAPTTTQAPTTTQA 186
             | |     |   | | |     |   |   | |       |   | | |    |
Db      81 PATAALNFAATAATPATAATPALIFAAAAATAATPATAALNFAATAATPATAATPALIFA 140

Qy     187 PTTTQAPTTITQA------ATTPA-ATTP----AATTPAATTPAATTPA---ATTPGVPA 232
           |   | |  | |       | ||| | ||    |||   | |||  |||   | |   ||
Db     141 ATAATAATPATPAFHFAATAATPATAATPALIFAATAATAATPA--TPAFHFAATAATPA 198

Qy     233 PTSAPVWPPICELLPNGCPA----DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYF 288
             : |              ||    :|     |          :   |       | |
Db     199 TAATPALIFAATAATAATPATAALNFAATAATPATAATPALIFAATAATAATPATAALNF 258

Qy     289 NPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRC 348
                 :||   |:   : | | |                   ||
Db     259 AATAATAATPATAACNFAATAATPATAAT--------------PA-------------- 290

Qy     349 DKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP 408
                                       |:                  :
Db     290 ---------------------------------LI----------------------FAA 294

Qy     409 TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHF-----SPAT 463
           |     |   | :|:     |  :    :     | | |    |:|     :|||
Db     295 TAATAATPATAAC--NFAATAATPATAATPALIFAATAATAATPATAALNFAATAATPAT 352

Qy     464 -----------QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAP 511
                     :  :| ||          |   : | |  | | |   | |||  | |
Db     353 AATPALIFAATAATAATPATAAL--------NFAATAATPATAATPALIFAATAATAATP 404

Qy     512 -----------STVVPPATPP---ATAAPVPPTTAIPTP--APTAAPTAAPTTAAPESPT 555
                   :|   ||||     |  | | | | |    | |||  | | |||
Db     405 ATAALNFAATAATAATPATPAFNFAATAATPATAATPALIFAATAATAATPATAALNFAA 464

Qy     556 TVTVPPTAAPTA---APTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVT 612
           |   | |||  |   | | |      |  :| |||   | |||  | | ||      |
Db     465 TAATPATAATPALIFAATAA------TAATAATAALHFAATAATAATPATAALNFAATAA 518

Qy     613 SPPTAA-PTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTA----APTAAPPTVAHA 667
           :| ||| |     |  |   | |||   || |   |    | ||    | ||| |  |
Db     519 TPATAATPALIFAATAATAATPATAAFNFAATA--ATAATPATAALNFAATAATPATAAT 576

Qy     668 PNTT-AAPVTTTSAPAT 683
           |    ||   | : |||
Db     577 PALIFAATAATAATPAT 593


RESULT  11
MUA1_XENLA
ID   MUA1_XENLA     STANDARD;      PRT;   400 AA.
AC   P10667;
DT   01-JUL-1989 (REL. 11, CREATED)
DT   01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   INTEGUMENTARY MUCIN A.1 PRECURSOR (FIM-A.1) (PREPROSPASMOLYSIN).
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88227968.
RA   HOFFMANN W.;
RL   J. BIOL. CHEM. 263:7686-7690(1988).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE; 90316191.
RA   HAUSER F., GERTZEN E.M., HOFFMANN W.;
RL   EXP. CELL RES. 189:157-162(1990).
CC   -!- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC       INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AND STORED EXCLUSIVELY IN MATURE
CC       MUCOUS GLANDS OF THE SKIN.
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED. CONSIST OF ABOUT 70% CARBOHYDRATE
CC       AND 30% PROTEIN.
CC   -!- SIMILARITY: CONTAINS 4 P-TYPE (TREFOIL) DOMAINS.
DR   EMBL; M19971; G214792; -.
DR   PIR; A28172; A28172.
DR   HSSP; P01359; 1PCP.
DR   PROSITE; PS00025; P_TREFOIL; 3.
KW   REPEAT; SIGNAL; AMPHIBIAN SKIN; GLYCOPROTEIN.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    400       INTEGUMENTARY MUCIN A.1.
FT   DOMAIN       22     63       P-TYPE 1.
FT   DOMAIN       73    114       P-TYPE 2.
FT   DOMAIN      127    252       14 X 9 AA TANDEM REPEATS.
FT   DOMAIN      272    299       7 X 4 AA REPEATS OF E-T-T-T.
FT   DOMAIN      304    342       P-TYPE 3.
FT   DOMAIN      352    393       P-TYPE 4.
FT   DISULFID     23     49       BY SIMILARITY.
FT   DISULFID     33     48       BY SIMILARITY.
FT   DISULFID     43     60       BY SIMILARITY.
FT   DISULFID     74    100       BY SIMILARITY.
FT   DISULFID     84     99       BY SIMILARITY.
FT   DISULFID     94    111       BY SIMILARITY.
FT   DISULFID    312    327       BY SIMILARITY.
FT   DISULFID    322    339       BY SIMILARITY.
FT   DISULFID    353    379       BY SIMILARITY.
FT   DISULFID    363    378       BY SIMILARITY.
FT   DISULFID    373    390       BY SIMILARITY.
FT   CARBOHYD     63     63       POTENTIAL.
SQ   SEQUENCE   400 AA;  42641 MW;  6DEA5D96 CRC32;


  Query Match             7.7%;  Score 331.5;  DB 1;  Length 400;
  Best Local Similarity   29.9%;  Pred. No. 1.8e-08;
  Matches  115;  Conservative   21;  Mismatches  142;  Indels  107;  Gaps   21;

Qy     434 ESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAG-CQVFECDSD---------- 483
           |:|| : : |    :|             |: |    |||  :  ||  |
Db      40 EADC-RAVGCCFDSSILN----------TKWCFYNATAGPIKKLECSGDPTKRIDCGFP 87

Qy     483 ----NQC-----------------TSTAAPTAAP---TAAPTAAPTAAPTAAPSTVVPP 517
               ||                     |   | ||   ||: ||  |  ||  | |   |
Db      88 RITEKQCILRGCCFDSSISGVKWCYARTVITTPAPDTTTASTTAETTTVPT-TPETTTVP 146

Qy     518 ATPPATAAP-VPPTTAIP-TPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPE 575
             ||  |  |  | || :| ||  |  ||   ||  | :| | |||  |   |  |||  ||
Db     147 TTPETTTVPTTPETTTVPTTPETTTVPTTPETTTVPTTPETTTVPTT--PE 204

Qy     576 IPITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPT 635
                 |: ||   |   |  ||   || ||  | | |:  ||  ||    | |  |  ||
Db     205 ----TTTVPTTPET---TTVPTTPETTTVPTTPETTTASTTAETTTVPTTPET--TTEPT 255
```

```
Qy    636 AAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLP 695
Db    256 TTPTTD--------TTPPTLPPT--PETTTETTTETTTETTTETTTETTTETTTAPPP-- 304

Qy    696 NDPINPCVE---ECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFN---PTTKTCDF-- 748
Db    304 -----ECAADRVDCGYSGITQADCE---------------GKGCIFDSTIPETKWCFYTE 343

Qy    748 ----ACNVGC-VRSNIQMSESYEGV 767
Db    344 AEAPARKAECTVDPSVRTDCGYPGI 368


RESULT  12
VG50_HSVI1
ID   VG50_HSVI1     STANDARD;      PRT;   670 AA.
AC   Q00130;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL GENE 50 PROTEIN.
GN   50.
OS   ICTALURID HERPESVIRUS 1 (CHANNEL CATFISH VIRUS) (CCV).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AUBURN 1;
RX   MEDLINE; 92087490.
RA   DAVISON A.J.;
RL   VIROLOGY 186:9-14(1992).
DR   EMBL; M75136; G331260; -.
DR   PIR; F36791; F36791.
KW   HYPOTHETICAL PROTEIN; REPEAT.
FT   REPEAT      143    158       1.
FT   REPEAT      171    186       2.
FT   REPEAT      200    214       3.
FT   REPEAT      215    233       4.
FT   REPEAT      234    252       5.
FT   REPEAT      253    268       6.
FT   REPEAT      279    293       7.
FT   REPEAT      294    309       8.
FT   REPEAT      320    334       9.
FT   REPEAT      335    349       10.
FT   REPEAT      362    376       11.
FT   REPEAT      377    391       12.
FT   REPEAT      392    406       13.
FT   REPEAT      407    421       14.
FT   REPEAT      422    436       15.
FT   REPEAT      437    452       16.
FT   REPEAT      464    477       17.
FT   REPEAT      478    493       18.
FT   REPEAT      504    517       19.
FT   REPEAT      518    531       20.
FT   REPEAT      532    545       21.
FT   REPEAT      546    559       22.
FT   REPEAT      560    573       23.
FT   REPEAT      574    587       24.
FT   REPEAT      588    601       25.
FT   REPEAT      602    615       26.
FT   REPEAT      616    629       27.
SQ   SEQUENCE   670 AA;  64174 MW;  FEC4F61D CRC32;


  Query Match             7.6%;  Score 330.5;  DB 1;  Length 670;
  Best Local Similarity   27.1%;  Pred. No. 3e-08;
  Matches  183;  Conservative   29;  Mismatches  262;  Indels  201;  Gaps   35;

Qy     96 PAETTQAPATTQA----------------PTTTQAPTT--TTQAPT------TTTQAPT 130
Db     72 PREPSSLPRTTPSLAHTTISKMISLGTRPRPTIPGVPTTIPNTDAPVDPGSVHTTARVVT 131

Qy    131 --TTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPT 188
Db    132 DITTKQTPTT---PATPAGANDTANITTATPAGANDTANITTATPAGANDTANITTATPA 188

Qy    189 TTQAPTTITQAATTPAATTPAATTPAATTPAAT------TPAATTPGVP----------A 232
Db    189 GAN-DTAVTTTSATPAGANDTAVTTTPATPAGANDTANDTVVTTTPAMPAGANDTANGTA 247

Qy    233 PTSAPVWPP-------ICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEG 285
Db    248 VTTTPAMPAGANDTANITTATPTGANDTANVTM----------------------PAG 283

Qy    286 LYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDI----GDLLDNGCPANFEIDWL 341
Db    284 ATDTVVTTTPAMPTGANDTANITTATP-AGANDTANVTMPAGATDTVVTTTPA------- 336

Qy    342 LPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA 401
Db    336 MPAGAN--------------------------------DTANVTKPAGSTDTVVTTTPA 362

Qy    402 CTGWYCPTEPIEWEPLPNGC--------PADFSIDHLLPHESDCGQYLQCVHGQTIARPC 453
Db    363 MPTGATDTVVTTTPAMPTGATDTVVTTTPA-------MPTGA-----TDTVVTTTPAKPA 410

Qy    454 PGNLHFSPATQSCESP-VTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPS 512
Db    411 GAN-----GTVVTTTPAMPAGA--------NDTVVTTAP-ATPAGANDTANVTKPTGATD 456

Qy    513 TVVPPAT--PPATAAPVPPTTAIPTPA--------PTAAP-TAAPTTAAPESPT-TVTVP 560
Db    457 TVVTTATVKPTGATGTVTTTTAKPTGANDTANVTKPTGATGTVTTTTAKPTGATGTVTV- 516

Qy    561 PTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPP----T 616
Db    516 ATAKPTGATGTVT-----TTTAKPTGANGTVTTT--TAKPTGATGTVTTTTAKPTGANGT 568

Qy    617 AAPTTAAPA-PNTTVTVPPTAAPTTAAPA-PNTTVTVPPTAAPTAAPPTVAHAPNTTAAP 674
Db    569 VTTTTAKPAGANGTVTT------TTAKPAGANGTVTT-TTAKPAGANGTVT---TTTAKP 618

Qy    675 -------VTTTSAPA 682
Db    619 AGANGTVTTTTAKPA 633


RESULT  13
VGP3_EBV
ID   VGP3_EBV       STANDARD;      PRT;   907 AA.
AC   P03200; P03201;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   ENVELOPE GLYCOPROTEIN GP340 (MEMBRANE ANTIGEN) (MA) (CONTAINS:
DE   GYLCOPROTEIN GP220).
GN   BLLF1.
OS   EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; GAMMAHERPESVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84270667.
RA   BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,
RA   GIBSON T.J., HATFULL G., HUDSON G.S., SATCHWELL S.C., SEGUIN C.,
RA   TUFFNELL P.S., BARRELL B.G.;
RL   NATURE 310:207-211(1984).
CC   -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
CC       B-CELLS.
CC   -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
CC       ENVELOPE.
DR   EMBL; V01555; G59109; -.
DR   EMBL; V01555; G832992; -.
DR   PIR; A03762; QQBE21.
DR   PIR; A03763; QQBE22.
DR   PIR; S33008; S33008.
KW   MEMBRANE; GLYCOPROTEIN; ANTIGEN; LATE PROTEIN; ALTERNATIVE SPLICING.
FT   VARSPLIC    502    698       MISSING (IN GP220).
```

```
FT   CARBOHYD     47     47       POTENTIAL.
FT   CARBOHYD     87     87       POTENTIAL.
FT   CARBOHYD    114    114       POTENTIAL.
FT   CARBOHYD    166    166       POTENTIAL.
FT   CARBOHYD    169    169       POTENTIAL.
FT   CARBOHYD    195    195       POTENTIAL.
FT   CARBOHYD    229    229       POTENTIAL.
FT   CARBOHYD    277    277       POTENTIAL.
FT   CARBOHYD    318    318       POTENTIAL.
FT   CARBOHYD    328    328       POTENTIAL.
FT   CARBOHYD    345    345       POTENTIAL.
FT   CARBOHYD    356    356       POTENTIAL.
FT   CARBOHYD    378    378       POTENTIAL.
FT   CARBOHYD    386    386       POTENTIAL.
FT   CARBOHYD    411    411       POTENTIAL.
FT   CARBOHYD    435    435       POTENTIAL.
FT   CARBOHYD    443    443       POTENTIAL.
FT   CARBOHYD    457    457       POTENTIAL.
FT   CARBOHYD    497    497       POTENTIAL.
FT   CARBOHYD    519    519       POTENTIAL.
FT   CARBOHYD    533    533       POTENTIAL.
FT   CARBOHYD    547    547       POTENTIAL.
FT   CARBOHYD    568    568       POTENTIAL.
FT   CARBOHYD    589    589       POTENTIAL.
FT   CARBOHYD    610    610       POTENTIAL.
FT   CARBOHYD    624    624       POTENTIAL.
FT   CARBOHYD    627    627       POTENTIAL.
FT   CARBOHYD    645    645       POTENTIAL.
FT   CARBOHYD    656    656       POTENTIAL.
FT   CARBOHYD    683    683       POTENTIAL.
FT   CARBOHYD    701    701       POTENTIAL.
FT   CARBOHYD    735    735       POTENTIAL.
FT   CARBOHYD    746    746       POTENTIAL.
FT   CARBOHYD    755    755       POTENTIAL.
FT   CARBOHYD    780    780       POTENTIAL.
FT   CARBOHYD    815    815       POTENTIAL.
FT   CARBOHYD    858    858       POTENTIAL.
FT   CARBOHYD    888    888       POTENTIAL.
SQ   SEQUENCE   907 AA;  94431 MW;  1EF4EBFD CRC32;
```

```
  Query Match           7.1%;  Score 309;  DB 1;  Length 907;
  Best Local Similarity  22.3%;  Pred. No. 3.1e-07;
  Matches  167;  Conservative   66;  Mismatches  221;  Indels  294;  Gaps   32;

Qy      43 WLLPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGC-TLPGPPAET-- 100
           |  |:  :    : |:: |    |  |       |::       ::|  | |     |
Db     352 WAWPNNTETD--FKCKWTLTSGTPSGC-ENISGAFASNRTFDITVSGLGTAPKTLIITRT 408

Qy     100 -TQAPATTQAPTTTQAPTTTTQAPTTTTQA---PTTTTQAPTTTQAPTTTQAPTTTQAPT 155
            | |  ||     ::|| :|| :||  |     | |||  |::|  ||   || :|  ||
Db     409 ATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPAST-GPT 467

Qy     156 TTTQAPTTTTQAPTTTTQAPTT-----------TQAP--TTTQAPTTTQAP--TTITQAA 200
           :|   |: | | ||:  :| |           ::||  |:: :| ||  |  |: | |
Db     468 VSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTSSTSPVTTPTPNATSPTPAV 527

Qy     201 TTPAATTPAAT--TPAATTPAATTPAATTP--GVPAPTSAPVWPPICELLPNGCPADFDI 256
           |||   || ||  ||| |||   || ||:|  |  :|||   |      ||
Db     528 TTP---TPNATSPTPAVTTP---TPNATSPTLGKTSPTSAVTTP-----TPN-------- 569

Qy     257 HLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEG 316
                                                 ||      |: ||   ||
Db     569 ------------------------------------ATSPTL----GKTSPTSAVT-- 585

Qy     317 NEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQ 376
                                    | |
Db     585 ------------------------TPTPN---------------------------- 590

Qy     377 CDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESD 436
               | |               |   ||  :   | ||
Db     590 --------ATSP-------------TLGKTSPTSAVT-TPTPNAT--------------- 613
```

```
Qy     437 CGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPT 496
                 | |:    |                          : :: |      :||   |
Db     613 ---------GPTVGETSP----------------------QANATNHTLGGTSPTPVVT 640

Qy     497 AAPTAAPTAA-----------------PTAAPSTVVPPATPPATAAPVPPTTAIPT--- 536
           : |  | :|                  |:: | |: |  :  :|:     |:| ||
Db     641 SQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSDNSTSHMPLLTSAHPTGGE 700

Qy     536 ----PAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVP-EIPITVTSAP-----T 585
               | :  |   :|::| :|   |    :  |  : |:  | |: :|  : |
Db     701 NITQVTPASISTHHVSTSSP-APRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNATSP 759

Qy     586 AAPTAAPTAAPTAAPTTAVPEIPT-----------TVTSPPT---AAPTTAAPAPNTTVT 631
            ||:   || ||   |       |           | | | |       ||  |  | |
Db     760 QAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRPRYNATTY 819

Qy     632 VPP------------TAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVT--- 677
           :||            |: | | | |   || ||||: |  :  ::      : | :|
Db     820 LPPSTSSKLRPRWTFTSPPVTTAQA---TVPVPPTSQPRFSNLSMLVLQWASLAVLTLLL 876

Qy     677 ------------TTSAPATTPEDDDID 691
                       :||   |||  || :
Db     877 LLVMADCAFRRNLSTSHTYTTPPYDDAE 904
```

```
RESULT  14
YJ9P_YEAST
ID   YJ9P_YEAST     STANDARD;      PRT;  1161 AA.
AC   P47179;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 118.4 KD PROTEIN IN BAT2-DAL5 INTERGENIC REGION
DE   PRECURSOR.
GN   YJR151C OR J2223.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   SCARCEZ T.;
RL   SUBMITTED (SEP-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE SRP1/TIP1 FAMILY OF STRESS-INDUCED
CC       PROTEINS.
DR   EMBL; Z49651; G1015903; -.
DR   PROSITE; PS00724; SRP1_TIP1; 1.
KW   HYPOTHETICAL PROTEIN; GLYCOPROTEIN; SIGNAL; GPI-ANCHOR.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1146       HYPOTHETICAL PROTEIN YJR151C.
FT   PROPEP     1147   1161       REMOVED IN MATURE FORM (POTENTIAL).
FT   LIPID      1146   1146       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   1161 AA;  118358 MW;  55125063 CRC32;
```

```
  Query Match           7.1%;  Score 309;  DB 1;  Length 1161;
  Best Local Similarity  19.5%;  Pred. No. 3.8e-07;
  Matches  120;  Conservative   70;  Mismatches  224;  Indels  200;  Gaps   13;

Qy      73 TEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTQAPTTT 132
           |  | |  |      :   |        ||   :||  |||:   || | : |:||   :||
Db     128 TTTKSSTSTTPTTTITSTTSTTSTTPTTSTTSTT--PTTSTTSTTPTTSTTSTTPTTSTT 185

Qy     133 TQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQA 192
           :  |||:   ||:  |||:   || | : |:||   :||:  |||:   |||:   ||:|
Db     186 STTPTTS---TTSTTPTTSTTSTTPTTSTTSTTSTTPTTSTTSTTPTTSTTSTTSQT 242

Qy     193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
           |  |   |:  :|||  :|   |:  :| |  :|    : ||
Db     243 STKSTTPTTSSTSTTPTTSTTPTTSTTSTAPTTSTTSTTSTTST---------------- 287

Qy     253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
```

```
Db    287 ----------------------------------------------------ISTAPT 292

Qy    313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372

Db    293 TST----------------------------------------------------TSSTF 300

Qy    373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLP 432

Db    301 STSSASASSVISTT--------------------ATTSTTFASLTTPATSTASTDHTTS 339

Qy    433 HESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPT 492

Db    340 SVSTTNAF--------------------TTSATTTTTSDTYI-SSSSPSQVTSSAEPT 376

Qy    493 AAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAI-PTPAPTAAPTAAPTTAAP 551

Db    377 TVSEVTSSVEPTR------SSQVTSSAEPTTVSEF--TSSVEPTRSSQVTSSAEPTTV-- 427

Qy    552 ESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTV 611

Db    427 -SEFTSSVEPTRSS----------QVTSSAEPTTVSEFTSSVEPTRSSQV------TSS 468

Qy    612 TSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTT 671

Db    469 AEPTTVSEFTSSVEPTRSSQVTSSAEPTTVSEFTSSVEPIRSSQVTSSAEPTTVSEVTSS 528

Qy    672 AAPVTTTSAPATTP 685

Db    529 VEPIRSSQVTTTEP 542
```

```
RESULT  15
CHI2_COCIM
ID   CHI2_COCIM     STANDARD;      PRT;   860 AA.
AC   P54197;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   ENDOCHITINASE 2 PRECURSOR (EC 3.2.1.14).
GN   CTS2.
OS   COCCIDIOIDES IMMITIS.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C735;
RX   MEDLINE; 96144270.
RA   PISHKO E.J., KIRKLAND T.N., COLE G.T.;
RL   GENE 167:173-177(1995).
CC   -!- FUNCTION: MAY BE ASSOCIATED WITH ENDOSPORULATION.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L41662; G1200192; -.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    860       ENDOCHITINASE 2.
FT   CARBOHYD     90     90       POTENTIAL.
SQ   SEQUENCE   860 AA;  91395 MW;  00E9C2B0 CRC32;
```

```
  Query Match             7.1%;  Score 308;  DB 1;  Length 860;
  Best Local Similarity   22.3%;  Pred. No. 3.3e-07;
  Matches  143;  Conservative   71;  Mismatches  220;  Indels  206;  Gaps   22;

Qy     93 PGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTT----TTQAPTTTQAPTTTQAP 148

Db    342 PDPP--TSTVTSTTSASTSTQTSSQSTTMETKTLSASTTPSSPSTVSPSSTM-QTTSTGS 398

Qy    149 TTTQAPTTTTQAPTTTTQAPTTTTQAPTTT--QAPTTTQAPTTTQAPTTITQAATTPAAT 206

Db    399 TSIETVTTRSQEPPSTTISTRSASTEPVTTRSQEPPSTTISTRSASTETVTTRSQEPPST 458

Qy    207 TPAATTPAATTPAAT----TPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPH 262

Db    459 T--ISTWSASTETSTSSQDSPSTTISTKSAPT---------------------------- 489

Qy    263 DKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDI 322

Db    489 ------------------------------------GTVTTRSQDLPSTTISTRSPETET 512

Qy    323 DIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIEL 382

Db    513 ETATTKSQGSPSI----------------------TLSTRSSSA-------------ET 536

Qy    383 VGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQ 442

Db    537 VSTRSQHSSSTTISTKS-------APTE--------TGTTSEHST--------------- 567

Qy    443 CVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAA 502

Db    567 ------------SMPVSTRSASTETVIT---RSQNSDSQSMTVSTRSPS---TESITTR 607

Qy    503 PTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPT 562

Db    608 SQGSPSETFSTKSVPVDTISTELPSQTPTTIITGTP-SDPVSAPTT---------TVPPN 657

Qy    563 AAPTAAPTTAVPEIPITVTS-APTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTT 621

Db    658 PTLTLAPSSSTTEDRTTITTIITTSYVTVCPTGFTTVTITYTTTYCPETASLTPTQPPIP 717

Qy    622 AAPAP-----NTTVTVPPTAAPTTAAPAPNT-TVTVPPTAAPTAAPPTVAHAPNTTAAPV 675

Db    718 GAPAPPPDGWTTIVTVCPQCAPT-----PTTVTLTVPTRSA--FLPARTETRPVVTVVPV 770

Qy    676 -----------------TTTSAPATTPEDDDIDPPLPND 697

Db    771 PENPIKNVKPSESGDFVTVTTAAPATVTKTLEYNNPVDSD 810
```

```
Search completed: July 18, 1999, 01:41:31
Job time: 2314 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 00:40:44 ; Search time 27.5 Seconds
(without alignments)
1073.409 Million cell updates/sec

Title: US-09-294-663-3
Perfect score: 4328
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 788

Scoring table: BLOSUM62

Searched: 116738 seqs, 37460341 residues

Database : PIR_58:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	636	14.7	3020	2	A43932	mucin 2 precursor,
2	440	10.2	873	2	A47283	calphotin - fruit
3	433	10.0	797	1	VGBEX1	glycoprotein X pre
4	432	10.0	662	2	A45155	mucin FIM-C.1 - Af
5	431	10.0	865	2	A47282	calcium-binding pr
6	397	9.2	1335	2	A35887	mucin precursor, p
7	397	9.2	1255	2	B35175	episialin B - huma
8	397	9.2	1295	2	A35886	polymorphic epithe
9	397	9.2	1264	2	A35175	episialin A - huma
10	393.5	9.1	1367	2	S48478	glucan 1,4-alpha-g
11	392	9.1	1188	2	S49915	extensin-like prot
12	381	8.8	294	2	A37232	mucin, tracheal (A
13	361.5	8.4	1118	2	A48292	mucin, tracheobron
14	354	8.2	822	2	A38420	antifreeze glycopr
15	351	8.1	474	2	S15921	protein TPX-VT3 -
16	345	8.0	232	2	A60095	larval glue protei
17	340.5	7.9	353	2	S36438	EPPT protein - hyd
18	340	7.9	279	2	S53363	mucin 5AC (clone J
19	339.5	7.8	901	2	A49227	sialidase - Actino
20	331.5	7.7	400	2	A28172	spasmolysin precur
21	330.5	7.6	670	2	F36791	hypothetical prote
22	328	7.6	379	2	S50125	larval'glue protei
23	324	7.5	528	2	I47141	gastric mucin (clo
24	322	7.4	216	2	I51920	mucin - rhesus mac
25	321.5	7.4	350	2	S22456	hydroxyproline-ric
26	314.5	7.3	303	2	S28264	hydroxyproline-ric
27	314	7.3	328	2	JQ0985	hydroxyproline-ric
28	309.5	7.2	327	2	S20074	promastigote surfa
29	309	7.1	907	1	QQBE21	membrane antigen g
30	309	7.1	1161	2	S57180	probable membrane
31	308.5	7.1	451	2	S71754	cellular hepatitis
32	308	7.1	860	2	JC4566	chitinase (EC 3.2.
33	307.5	7.1	307	1	GSFF3	salivary glue prot
34	306.5	7.1	1630	2	A53577	ascites sialoglyco
35	306	7.1	267	2	S08314	cell wall glycopro
36	305.5	7.1	700	2	A54641	interspersed repea
37	303.5	7.0	377	2	A48018	mucin 7 precursor,
38	303	7.0	447	2	A39321	mucin - rat (fragm
39	301.5	7.0	886	2	S29605	glycoprotein 350/2
40	301	7.0	322	2	A53715	apomucin precursor
41	298	6.9	283	2	S13383	hydroxyproline-ric
42	297	6.9	620	2	S06733	hydroxyproline-ric
43	296.5	6.9	376	2	S71558	probable cell wall
44	294.5	6.8	3149	1	QQBE8	BPLF1 protein - hu
45	292	6.7	477	2	S53362	mucin 5AC (clone J

ALIGNMENTS

RESULT 1
A43932
mucin 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin SMUC-41
C;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 20-Mar-1998
C;Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A;Reference number: A49963
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU1>
A;Cross-references: GB:L21998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A;Reference number: A45106; MUID:93016075
A;Accession: A45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 626-1895 <GU2>
A;Cross-references: GB:M94131; NID:g186395; PID:g186396
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)
A;Accession: B45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2037-3020 <GU3>
A;Cross-references: GB:M94132; NID:g186397; PID:g186398
A;Experimental source: colon
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A;Reference number: A43932; MUID:91358717
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350,'L',1352-1411,'S',1413-1448,'P',1450-1503,'T',1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:g188863; PID:g188864
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A;Reference number: A33532; MUID:89197956
A;Accession: B33532
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:g188873; PID:g188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
J. Clin. Invest. 87, 77-82, 1991
A;Title: Human bronchus and intestine express the same mucin gene.
A;Reference number: A61257
A;Accession: A61257
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>
A;Experimental source: bronchus
R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstn
Biochem. Biophys. Res. Commun. 183, 821-828, 1992

A;Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t
A;Reference number: PQ0328; MUID:92198477
A;Accession: PQ0328
A;Molecule type: mRNA
A;Residues: 2328-2468 <XUG>
A;Cross-references: GB:M86523
A;Experimental source: small intestine
A;Accession: PQ0329
A;Molecule type: protein
A;Residues: 2328-2342,'K',2344-2354 <XUG1>
C;Genetics:
A;Gene: GDB:MUC2
A;Cross-references: GDB:120203; OMIM:158370
A;Map position: 11p15.5-11p15.5
C;Superfamily: von Willebrand factor type C repeat homology
C;Keywords: glycoprotein; intestine; tandem repeat
F;2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 14.7%; Score 636; DB 2; Length 3020;
Best Local Similarity 25.6%; Pred. No. 1.1e-24;
Matches 201; Conservative 41; Mismatches 213; Indels 330; Gaps 24;

```
Qy      29 LHEPHP-------DCPPAEQHWLLPHEYDCTKF------------YYCEYGLKFIAPRD 68
           ::| ||       |  | :    | : :|                 |:  : ||   :
Db    1309 INEDHPSSGSDDGDREPFDGVCGAPEDIECRSVKDPHLSLEQHGQKVQCDVSVGFICKNE 1368

Qy      69 CAPGT-------EFKFSAQTCVHAALAGC-TLPGPPAETTQAPATTQAPTTTTQAPTTTTQ 120
             |        ::|    |    :  | | | ||  |   | ||   |||  ||||
Db    1369 DQFGNGPFGLCYDYKIRVNCC--WPMDKCITTPSPPTTTPSPPPTT---TTTLPPTTTPS 1423

Qy     121 APTTTTQAPTTTTQAPTTTQAP----TTTQAPTTTQAP---TTTTQAPTTTTQAPTTTTQ 173
            |||     ||||  |||| :|    |||  |||| :|   ||||  ||||   ||||
Db    1424 PPTT-----TTTTPPPTTTPSPPITTTTTPLPTTTPSPPISTTTTPPPTTTPSPPTTTPS 1478

Qy     174 APTTTQAP----TTTQAPTTTQAPTTITQAATTPAATT--PAATTPAATTPAATTPAATT 227
            |||| :|    |||  |||| :|  :|   | ||:||  |  |||:  |   |||  ||
Db    1479 PPTTTPSPPTTTTTTPPPTTTPSP-PMTTPITPPASTTTLPPTTTPSPPTTTTTTPPPTT 1537

Qy     228 PGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLY 287
              | ||: |: ||
Db    1538 TPSP-PTTTPITPP---------------------------------------------- 1551

Qy     288 FNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNR 347

Db    1551 ------------------------------------------------------------ 1551

Qy     348 CDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYC 407

Db    1551 ------------------------------------------------------------ 1551

Qy     408 PTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCE 467

Db    1551 ------------------------------------------------------------ 1551

Qy     468 SPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPS-----TVVPP----A 518
                          |:|  ||  |:  ||   |  ||  ||      |  ||
Db    1551 ---------------TSTTTLPPTTTPSPPPTTTTPPPTTTPSPPTTTTPSPPTITTT 1594

Qy     519 TPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPI 578
           ||| |  | ||||   || ||  |:   ||       :| |:|||  |:  |||      |
Db    1595 TPPPTTTPSPPTTTTTTPPPTTTPSPPTTTPITPPTSTTTLPPTTTPSPPPTTTTTPPPT 1654

Qy     579 TVTSAPTAAPTAAPTAAPTAAPTTAVPEIP-TTVTSPPTAAPTTAAPAPNTTVTVPPTAA 637
           |  | ||    : |   |  | | |  |  | ||  ||||   ||  |:| || : | |
Db    1655 TTPSPPTTTTPSPPITTTTTPPPTTTPSSPITTTPSPPTTTMTT--PSPTTTPSSPIT-- 1711

Qy     638 PTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNT--TAAPVTTTSAP-ATTPEDDDIDPPL 694
            ||  |:  || : |||   | :| |    | |  |  | ||||:|  |||    | ||
Db    1711 -TTTTPSSTTTPSPPPTTMTTPSPTTTPSPPTTTMTTLPPTTTSSPLTTTPLPPSITPPT 1769

Qy     695 -----PNDPINPCVEECNVLPWAHADCDKYWVCDGNNQVL-VVCSEGLQFNPTTKTCDFA 748
                 |  |||  ||   |  :    :    |: :::  ||  |           :|
```

Db 1770 FSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDTELIGDVCGPG----------WA 1818

Qy 749 CNVGC 753
|: |
Db 1819 ANISC 1823

RESULT 2
A47283
calphotin - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 10-Sep-1997
C;Accession: A47283
R;Ballinger, D.G.; Xue, N.; Harshman, K.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 1536-1540, 1993
A;Title: A Drosophila photoreceptor cell-specific protein, calphotin, binds calcium a
A;Reference number: A47283; MUID:93165730
A;Contents: photoreceptor cells
A;Accession: A47283
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-873 <BAL>
A;Cross-references: GB:L05080; NID:g157071; PID:g157072
A;Note: sequence extracted from NCBI backbone (NCBIN:124958, NCBIP:124959)

Query Match 10.2%; Score 440; DB 2; Length 873;
Best Local Similarity 25.9%; Pred. No. 2.1e-15;
Matches 185; Conservative 42; Mismatches 248; Indels 240; Gaps 32;

```
Qy      91 TLPGPPAETTQAPAT---TQAPTTTQAPTTTTQAPTTTTQAP------------------ 130
           |:| | :    || |     ||    :|     ||     ||
Db      5 TIPSPVSAPVAAPVTPSAVAAPVQVVSPAAVAVAPAVVAPAPAAPTAVTPVAPPPTLASV 64

Qy     130 ---TTTTQAP----TTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPT 182
             | |  ||      : ||  : ||    |||    :| :|   |    | |    ||
Db      65 QPATVTVPAPAPIAAASVAPVASVAPPVVAAPTPPAASPVST--PPVAVAQIPVAVSAPV 122

Qy     183 TTQAPTTTQAPTTITQ-AATTPAATTP----AATTPAATTPAATTPAATTPGVPAPTSAP 237
             ||     || :       |   ||    :| |||| ||  :   || | ||| |: |
Db     123 ---APPVAATPTPVAPIPVAAPVIATPPVAASAPTPAAVTPVVSPVIATPPVVPANTTVP 179

Qy     238 VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDS 297
           |  |:   :|   |                                  | :   :
Db     180 VAAPVA-AVPAAVPV---------------------------------VAPVLAPAVA 203

Qy     298 PANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHG 357
           ||      |    ||| |      | :  : :   |       |:|  :
Db     204 PAVAPVVAETPAPPPVAE------IPVATIPECVAP-------LIPEVS---------- 240

Qy     358 NLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPL 417
                                   |  | |   :| | |   |       ||
Db     240 -------------------------VVATKPLAAAEPVVVAPPA-------TE------T 261

Qy     418 PNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVF 477
           |   ||  |    || |             |      :| : | | || |
Db     262 PVVAPAAAS-----PHVS--------------VAPAVETAVVAPVSASTEPPVAA----- 298

Qy     478 ECDSDNQCTSTAAPTAAPTAAPTAA-------------PTAAPTAAPSTVVPP------- 518
           |  | ||   |  ||  |               || ||   | |: ||
Db     298 -------ATLTTAP-ETPALAPVVAESQVAANTVVATPPTPAP--EPETIAPPVVAETPE 347

Qy     518 --------ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVT-----VPPTAA 564
                   |||     |||  |   |||  | |  | | |    |          :||  |
Db     348 VASVAVAETTPPV----VPPVAAESIPAPVVATTPVPATLAVTDPDVTASAVPELPPVIA 403

Qy     565 PTAAPTTAVPEIPITVTSAPTAAPTAA------------PTAAPTAAPTTA----VPEIP 608
           |:  | :|| | |: : : |    | ||             | || :|| |     :||:
Db     404 PSPVP-SAVAETPVDL-APPVLPPVAAEPVPAVVAEETPETPAPASAPVTIAALDIPEVA 461

Qy     609 TTVTSP---PTAAPTTAAP---APNTTVTVPPTAAPTTAAPAPNTTVTVPPTAA---PTA 659
             : :|   |  ||: |||    | || :|| | ||  | |     |:|   ||   | |
Db     462 PVIAAPSDAPAEAPSAAAPIVSTPPTTASVPETTAPPAAVPTEPIDVSVLSEAAIETPVA 521
```

```
Qy    660 APPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHA 714
Db    522 PPVEV-----TTEVAVADV-APPEAAEDLIIEPVEPPAPIPDLLEQTTSVPAVEA 570

RESULT   3
VGBEX1
glycoprotein X precursor - equine herpesvirus 1 (strain Ab4p)
C;Species: equine herpesvirus 1
A;Note: host Equus caballus (domestic horse)
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 05-Sep-1997
C;Accession: H36802
R;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
submitted to GenBank, March 1992
A;Description: The DNA sequence of equine herpesvirus-1.
A;Reference number: A36805
A;Accession: H36802
A;Molecule type: DNA
A;Residues: 1-797 <TEL>
A;Cross-references: GB:M86664; NID:g330791; PID:g330862
R;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
Virology 189, 304-316, 1992
A;Title: The DNA sequence of equine herpesvirus-1.
A;Reference number: A41831; MUID:92295566
A;Contents: annotation; possible protein-coding frames
A;Note: neither amino acid nor nucleotide sequence is given
C;Genetics:
A;Gene: 71
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homo
C;Keywords: glycoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-797/Product: glycoprotein X #status predicted <MAT>
F;23-465/Region: serine/threonine-rich
F;489-797/Domain: equine herpesvirus 1 glycoprotein homology <EHG>
F;766-790/Domain: transmembrane #status predicted <TMN>
F;590/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match          10.0%;  Score 433;  DB 1;  Length 797;
  Best Local Similarity  25.1%;  Pred. No. 4.4e-15;
  Matches  181;  Conservative   40;  Mismatches  167;  Indels  334;  Gaps   21;

Qy     78 SAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTQAPTTTQAPTT--TTQA 135
Db     77 STQSSSTAATSS-SAPSTASSTTSIPTSTSTETTTTTPTASTTTPTTTTAAPTTAATTTA 135

Qy    136 PTT---TQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTT--------------------TT 172
Db    136 VTTAASTSAETTTATATATSTPTTTTPTSTTTTTATTTVPTTASTTTDTTTAATTTAATT 195

Qy    173 QAPTTTQAPT---TTQAPTTTQAPTT------ITQAATTPAATTPAATTPAATTPAATTP 223
Db    196 TAATTTAATTTAATTTAATTTAATTTAATTSSATTAATTTAATTTAATTTAATTTAATTT 255

Qy    224 AATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCP 283
Db    256 AATTTG--SPTS---------------------------------------------------- 266

Qy    284 EGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLP 343
Db    266 ------------------------------------------------------------ 266

Qy    344 HGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACT 403
Db    266 ---------------------------------------GSTSTTGASTSTP------- 279

Qy    404 GWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPAT 463
Db    279 ------------------------------------------------------SAST 282

Qy    464 QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPA- 523
Db    283 ATSATP----------------TSTSTSAAATTSTPT--PTSAATSAESTTEAPTSTPTT 324

Qy    523 -TAAPVPPTTAIPTPAPT---AAPTAAPTTA-------APESPT--TVTVPPTAAPTAAP 569
Db    325 DTTTPSEATTATTSPESTTVSASTTSATTTAFTTESHTSPDSSTGSTSTAEPSSTFTLTP 384

Qy    570 TTAVPEI-PITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNT 628
Db    385 STATPSTDQFTGSSASTESDSTDSSTVPTTGTESITESSSTTEASTNLGSSTYESTEALE 444

Qy    629 TVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAPP-------------TVAHA------- 668
Db    445 TPDGNTTSGNTTPSPSPRT-----PSFADTQQTPDNGVSTQHTTINDHTTANAQKHAGHH 499

Qy    668 --------------PNTTAAPVTTTSAPATTPEDD-----------DIDPPLPNDPINPC 702
Db    500 RGRAGGRRGSPQGGSHTTPHPDRLTPSPDDTYDDDTNHPNGRNNSIEIVPQLP--PDRPI 557

Qy    703 VE 704
Db    558 IE 559

RESULT   4
A45155
mucin FIM-C.1 - African clawed frog (fragment)
C;Species: Xenopus laevis (African clawed frog)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Sep-1997
C;Accession: A45155
R;Hauser, F.; Hoffmann, W.
J. Biol. Chem. 267, 24620-24624, 1992
A;Title: P-domains as shuffled cysteine-rich modules in integumentary mucin C.1 (FIM-
A;Reference number: A45155
A;Accession: A45155
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-662 <HAU>
A;Cross-references: GB:L02115; NID:g214147; PID:g951460
C;Superfamily: trefoil homology
F;162-202/Domain: trefoil homology <TRF1>
F;307-347/Domain: trefoil homology <TRF2>
F;354-394/Domain: trefoil homology <TRF3>
F;526-566/Domain: trefoil homology <TRF4>
F;573-613/Domain: trefoil homology <TRF5>
F;621-661/Domain: trefoil homology <TRF6>

  Query Match          10.0%;  Score 432;  DB 2;  Length 662;
  Best Local Similarity  22.6%;  Pred. No. 4.2e-15;
  Matches  191;  Conservative   27;  Mismatches  229;  Indels  400;  Gaps   27;

Qy     65 APRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTT----QAPTT--- 118
Db     73 APATTAAGKAPTTAAATAPTTAAAG----APTTATGKAPATAAAPVPTTAASKAPTTAAA 128

Qy    118 ------TTQAPTTTTQA---------------------------------------- 129
Db    129 ATHSTAAAAAPTTAASAAKSKERSTSSSSEEEHCHVKPSKREMCGSKGITKKQCKKKNCC 188

Qy    129 ------------------------PTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTT 164
Db    189 FDPKGHGGIHCFHRKPKGHSHEEHTTTTTKAPTTIQIATTTTTPTTTTTTTKATPTTTTT 248

Qy    165 TQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPA 224
Db    249 TKATPTTT---TTTKATTTTTTPTTT---TTTTKATTTPTTTT--TTTPTTTTTKATTTT 300

Qy    225 ATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPE 284
Db    301 TTTSG---------------------------------------------------- 306

Qy    285 GLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPH 344
```

```
Db    306 -----------------ECKMEPS----------------------------------- 313

Qy    345 GNRCDKYYQCVHGNLVERRCGA-GTHFSFELQQ---CDHIELVGCTLPGGESEEVDVDED 400
Db    313 -----KREDCGYSGITESQCRTKGCCFDSSIPQTKWCFY------TL----SQVAD---- 354

Qy    401 ACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFS 460
Db    354 ------CKVEPSQ--------------------RVDCG---------------------- 366

Qy    461 PATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATP 520
Db    366 ----------FRGITADQCRQKNCCFDSSI-----SGTKWCFYSTSQVAATKTTTTPTT- 410

Qy    521 PATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITV 580
Db    410 -----TTTPTTTTTTKATTTTPTTTTTT-----PTTTTT---------------------- 439

Qy    581 TSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTT 640
Db    439 --------TTTTTKATTTTPTTTTP--TTTTTKATTTTPTTTTTTPTTTTTKATTTTPTT 488

Qy    641 AAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPIN 700
Db    489 TTTTPTTTTTKATTTTPTTTTTTT----TTTKATTTTTSGECKMEPSKRADCGYPGITES 544

Qy    701 PC----------------------VEECNVLPWAHADC---------------------- 717
Db    545 QCRSKGCCFDSSIPQTKWCFYSLPQVADCKVAPSSRVDCGFGGITADQCRQRNCCFDSSI 604

Qy    717 -DKYWVCDGNNQVLVVCSEGLQFNPTTKTCD---------FACNVGCVRSNIQMSESYEG 766
Db    605 SGTKWCFYSTSQGNAMCS-----GPPTKRRDCGYPGISSSVCINRGCCWDNSVMN----- 655

Qy    767 VQVFIPW 773
Db    655 ----VPW 657


RESULT   5
A47282
calcium-binding protein calphotin - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C;Accession: A47282
R;Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 1531-1535, 1993
A;Title: Calphotin: a Drosophila photoreceptor cell calcium-binding protein.
A;Reference number: A47282; MUID:93165729
A;Accession: A47282
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-865 <MAR>
A;Cross-references: GB:L02111; NID:g157031; PID:g157032
A;Experimental source: photoreceptor cells
A;Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBIP:124956)
C;Genetics:
A;Gene: FlyBase:Cpn
A;Cross-references: FlyBase:FBgn0010218
C;Keywords: calcium binding


  Query Match             10.0%;  Score 431;  DB 2;  Length 865;
  Best Local Similarity   25.8%;  Pred. No. 5.9e-15;
  Matches  183;  Conservative   43;  Mismatches  248;  Indels  234;  Gaps   32;

Qy     91 TLPGPPAETTQAPAT---TQAPTTTQAPTTTTQAPT-------------TTTQAPTTTT 133
Db      5 TIPSPVSAPVAAPVTPSAVAAPVQVVSPAAVAPAPAAPIAVTPVAPPPTLASVQPATVTI 64

Qy    134 QAPTTTQA----PTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTT 189
Db     65 PAPAPIAAASVTPVASVAPPVVAAPTPPAASPVST---PVAVAQIPVAVSAPV---APPV 118

Qy    190 TQAPTTITQ-AATTPAATTP----AATTPAATTPAATTPAATTPGVPAPTSAPVWPPICE 244
Db    119 AATPTPVVQIPVAAPVIATPPVAASAPTPAAVTPVISPVIASPPVVPANTTVPVAAPVA- 178

Qy    245 LLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECD 304
Db    178 AVPAAVPV-----------------------------------VAPVLAPAVAPAVAPVV 202

Qy    305 GEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRC 364
Db    203 AETPAPPPVAE------IPVATIPECVAP-------LIPEVS------------------ 232

Qy    365 GAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPAD 424
Db    232 ------------------VVATKPLAAAEPVVVAPPA-------TE------TPVVAPAA 260

Qy    425 FSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQ 484
Db    261 AS-----PHVS--------------VAPAVETAVVAPVSASTEPPVAA------------ 290

Qy    485 CTSTAAPTAAPTAAPTAA-------------PTAAPTAAPSTVVPP-------------- 518
Db    290 ATLTTAP-ETPALAPVVAESQVAANTVVATPPTPAP--EPETIAPPVVAETPEVASVAVA 346

Qy    518 -ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVT-----VPPTAAPTAAPTT 571
Db    347 ETTPPV----VPPVAAESIPAPVVATTPVPATLAVTDPDVTASAVPELPPVIAPSPVP-S 401

Qy    572 AVPEIPITVTSAPTAAPTAA------------PTAAPTAAPTTA----VPEIPTTVTSP- 615
Db    402 AVAETPVDL-APPVLPPVAAEPVPAVVAEETPETPAPASAPVTIAALDIPEVAPVIAAPS 460

Qy    615 --PTAAPTTAAP---APNTTVTVPPTAAPTTAAPAPNTTVTVPPTAA---PTAAPPTVAH 666
Db    461 DAPAEAPSAAAPIVSTPPTTASVPETTAPPAAVPTEPIDVSVLSEAAIETPVAPPVEV-- 519

Qy    667 APNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHA 714
Db    519 ---TTEVAVADV-APPEAAADLIIEPVEPPAPIPDLLEQTTSVPAVEA 562


RESULT   6
A35887
mucin precursor, pancreatic - human
C;Species: Homo sapiens (man)
C;Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 03-Feb-1994
C;Accession: A35887
R;Lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A;Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A;Reference number: A35887; MUID:90368716
A;Accession: A35887
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1335 <LAN>
A;Cross-references: GB:J05582
C;Superfamily: polymorphic epithelial mucin


  Query Match              9.2%;  Score 397;  DB 2;  Length 1335;
  Best Local Similarity   26.1%;  Pred. No. 4.1e-13;
  Matches  199;  Conservative   42;  Mismatches  288;  Indels  234;  Gaps   34;

Qy     10 ALGLVA--ARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYGLKFIAPR 67
Db    107 ALGSTTPPAHDVTSAPDNKPA---PGSTAPPA------------------HGVTSAPDT 144

Qy     68 DCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTT-------TQ 120
Db    145 RPAPGS----TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 200

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTTQAPTTTQAPTTTQ 179
```

```
Db    201 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 260

Qy    180 APTTTQAPTTTQAPT-TITQAATT---PAATTPAA---TTPAATTPAATTPAATTPGVPA 232
Db    261 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA---PGSTAPPAHG 317

Qy    233 PTSAP--------VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTF--EQRC 282
Db    318 VTSAPDTRPAPGSTAPP-----------------------------AHGVTSAPDTRP 346

Qy    283 PEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLL 342
Db    347 APGSTAPPAHGVTSAPDTRPAPG--STAPP----------------------------- 375

Qy    343 PHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDAC 402
Db    375 ------------AHG---------------VTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 407

Qy    403 TGWYCP-----TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCP 454
Db    408 PGSTAPPAHGVTSAPDTRPAP------------------GSTAPPAHGVTSAPDTRPAP 448

Qy    455 GNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTV 514
Db    449 GST-APPAHGVTSAP----------DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 497

Qy    515 V--PPATPPATAAPVPP---TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAP 569
Db    498 VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 557

Qy    570 TTAVPEI-PITVTSAP-----TAAPTAAPTAAPTAAPTTAVPEIPTT-----VTSPP--- 616
Db    558 VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 617

Qy    616 -TAAPTTAAPAPNTTVTVPP----TAAPTTAAPAPNTTVTVPPTAAPTAAPPT------- 664
Db    618 VTSAPDT-RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPAHGVTSAPDTRPAPGST 671

Qy    664 -----VAHAPNTTAAPVTT-------TSAPATTPEDDDIDPP 693
Db    672 APPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 714
```

RESULT 7
B35175
episialin B - human
C;Species: Homo sapiens (man)
C;Date: 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change 23-Feb-1997
C;Accession: B35175
R;Ligtenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A;Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene enc
A;Reference number: A35175; MUID:90202794
A;Accession: B35175
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1255 <LIG>
A;Cross-references: GB:M34088; GB:J05288
C;Superfamily: polymorphic epithelial mucin
C;Keywords: cytoskeleton; transmembrane protein

Query Match 9.2%; Score 397; DB 2; Length 1255;
Best Local Similarity 26.1%; Pred. No. 3.9e-13;
Matches 199; Conservative 42; Mismatches 288; Indels 234; Gaps 34;

```
Qy     10 ALGLVA--ARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYGLKFIAPR 67
Db    107 ALGSTTPPAHDVTSAPDNKPA---PGSTAPPA------------------HGVTSAPDT 144

Qy     68 DCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTT-------TQ 120
```

```
Db    145 RPAPGS----TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 200

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTTQAPTTTTQAPTTTQ 179
Db    201 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 260

Qy    180 APTTTQAPTTTQAPT-TITQAATT---PAATTPAA---TTPAATTPAATTPAATTPGVPA 232
Db    261 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA---PGSTAPPAHG 317

Qy    233 PTSAP--------VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTF--EQRC 282
Db    318 VTSAPDTRPAPGSTAPP-----------------------------AHGVTSAPDTRP 346

Qy    283 PEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLL 342
Db    347 APGSTAPPAHGVTSAPDTRPAPG--STAPP----------------------------- 375

Qy    343 PHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDAC 402
Db    375 ------------AHG---------------VTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 407

Qy    403 TGWYCP-----TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCP 454
Db    408 PGSTAPPAHGVTSAPDTRPAP------------------GSTAPPAHGVTSAPDTRPAP 448

Qy    455 GNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTV 514
Db    449 GST-APPAHGVTSAP----------DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 497

Qy    515 V--PPATPPATAAPVPP---TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAP 569
Db    498 VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 557

Qy    570 TTAVPEI-PITVTSAP-----TAAPTAAPTAAPTAAPTTAVPEIPTT-----VTSPP--- 616
Db    558 VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 617

Qy    616 -TAAPTTAAPAPNTTVTVPP----TAAPTTAAPAPNTTVTVPPTAAPTAAPPT------- 664
Db    618 VTSAPDT-RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPAHGVTSAPDTRPAPGST 671

Qy    664 ------VAHAPNTTAAPVTT-------TSAPATTPEDDDIDPP 693
Db    672 APPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 714
```

RESULT 8
A35886
polymorphic epithelial mucin precursor - human
C;Species: Homo sapiens (man)
C;Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 03-Feb-1994
C;Accession: A35886
R;Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burc
J. Biol. Chem. 265, 15286-15293, 1990
A;Title: Molecular cloning and expression of human tumor-associated polymorphic epith
A;Reference number: A35886; MUID:90368715
A;Accession: A35886
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1295 <GEN>
A;Cross-references: GB:J05581
C;Superfamily: polymorphic epithelial mucin

Query Match 9.2%; Score 397; DB 2; Length 1295;
Best Local Similarity 26.1%; Pred. No. 4e-13;
Matches 199; Conservative 42; Mismatches 288; Indels 234; Gaps 34;

```
Qy     10 ALGLVA--ARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYGLKFIAPR 67
Db    107 ALGSTTPPAHDVTSAPDNKPA---PGSTAPPA------------------HGVTSAPDT 144
```

```
Qy      68 DCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTT-------TQ 120
Db     145 RPAPGS----TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 200

Qy     121 APTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTTQAPTTTTQAPTTTQ 179
Db     201 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 260

Qy     180 APTTTQAPTTTQAPT-TITQAATT---PAATTPAA---TTPAATTPAATTPAATTPGVPA 232
Db     261 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA---PGSTAPPAHG 317

Qy     233 PTSAP--------VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTF--EQRC 282
Db     318 VTSAPDTRPAPGSTAPP----------------------------AHGVTSAPDTRP 346

Qy     283 PEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLL 342
Db     347 APGSTAPPAHGVTSAPDTRPAPG--STAPP---------------------------- 375

Qy     343 PHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDAC 402
Db     375 ------------AHG--------------VTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 407

Qy     403 TGWYCP-----TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCP 454
Db     408 PGSTAPPAHGVTSAPDTRPAP----------------GSTAPPAHGVTSAPDTRPAP 448

Qy     455 GNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTV 514
Db     449 GST-APPAHGVTSAP----------DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 497

Qy     515 V--PPATPPATAAPVPP---TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAP 569
Db     498 VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 557

Qy     570 TTAVPEI-PITVTSAP-----TAAPTAAPTAAPTAAPTTAVPEIPTT-----VTSPP--- 616
Db     558 VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 617

Qy     616 -TAAPTTAAPAPNTTVTVPP----TAAPTTAAPAPNTTVTVPPTAAPTAAPPT------- 664
Db     618 VTSAPDT-RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPAHGVTSAPDTRPAPGST 671

Qy     664 ------VAHAPNTTAAPVTT-------TSAPATTPEDDDIDPP 693
Db     672 APPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 714
```

RESULT 9
A35175
episialin A - human
C;Species: Homo sapiens (man)
C;Date: 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change 23-Feb-1997
C;Accession: A35175
R;Ligtenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A;Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene enc
A;Reference number: A35175; MUID:90202794
A;Accession: A35175
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1264 <LIG>
A;Cross-references: GB:M32738; GB:M34088; GB:J05288
C;Superfamily: polymorphic epithelial mucin
C;Keywords: cytoskeleton; transmembrane protein

Query Match 9.2%; Score 397; DB 2; Length 1264;
Best Local Similarity 26.1%; Pred. No. 3.9e-13;
Matches 199; Conservative 42; Mismatches 288; Indels 234; Gaps 34;

```
Qy      10 ALGLVA--ARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYGLKFIAPR 67
Db     116 ALGSTTPPAHDVTSAPDNKPA---PGSTAPPA----------------HGVTSAPDT 153

Qy      68 DCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTT-------TQ 120
Db     154 RPAPGS----TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 209

Qy     121 APTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTTQAPTTTTQAPTTTQ 179
Db     210 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS 269

Qy     180 APTTTQAPTTTQAPT-TITQAATT---PAATTPAA---TTPAATTPAATTPAATTPGVPA 232
Db     270 APDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA---PGSTAPPAHG 326

Qy     233 PTSAP--------VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTF--EQRC 282
Db     327 VTSAPDTRPAPGSTAPP----------------------------AHGVTSAPDTRP 355

Qy     283 PEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLL 342
Db     356 APGSTAPPAHGVTSAPDTRPAPG--STAPP---------------------------- 384

Qy     343 PHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDAC 402
Db     384 ------------AHG--------------VTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 416

Qy     403 TGWYCP-----TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCP 454
Db     417 PGSTAPPAHGVTSAPDTRPAP----------------GSTAPPAHGVTSAPDTRPAP 457

Qy     455 GNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPSTV 514
Db     458 GST-APPAHGVTSAP----------DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 506

Qy     515 V--PPATPPATAAPVPP---TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAP 569
Db     507 VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 566

Qy     570 TTAVPEI-PITVTSAP-----TAAPTAAPTAAPTAAPTTAVPEIPTT-----VTSPP--- 616
Db     567 VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHG 626

Qy     616 -TAAPTTAAPAPNTTVTVPP----TAAPTTAAPAPNTTVTVPPTAAPTAAPPT------- 664
Db     627 VTSAPDT-RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPAHGVTSAPDTRPAPGST 680

Qy     664 ------VAHAPNTTAAPVTT-------TSAPATTPEDDDIDPP 693
Db     681 APPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 723
```

RESULT 10
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR01
C;Species: Saccharomyces cerevisiae
C;Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 05-Dec-1997
C;Accession: S48478; A26877; B26877; S27281; JC6123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: GB:Z47047; EMBL:Z38061; MIPS:YIR019c; NID:g603997; PID:g763364
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>

```
A;Cross-references: EMBL:M16164; NID:g172522; PID:g172525
A;Accession: B26877
A;Molecule type: DNA
A;Residues: 762-1331 <YA2>
A;Cross-references: EMBL:M16165; NID:g172523; PID:g172526
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A;Reference number: S27281
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A;Reference number: JC6123
A;Accession: JC6123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: GB:U30626; NID:g1304386; PID:g1304387
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2
A;Cross-references: MIPS:YIR019c; SGD:S0001458
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F;5-21/Domain: transmembrane #status predicted <TM1>
F;1350-1366/Domain: transmembrane #status predicted <TM2>

  Query Match             9.1%;  Score 393.5;  DB 2;  Length 1367;
  Best Local Similarity   19.8%;  Pred. No. 6.3e-13;
  Matches  150;  Conservative  120;  Mismatches  229;  Indels  259;  Gaps   28;

Qy      32 PHPDCPPAEQHWLLPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCT 91
           || |  |            |||        |    : |    |        :      :
Db     288 PHHDTTP-----------CTK-------KKTTTSKTCTKKTTTPVPTPSSSTTESSSAP 328

Qy      92 LPGPPAETTQ---APATTQAPTTTQAPTTTTQAPTT------TTQAPTTTTQAPTTTQAP 142
           :| | : ||:   || |:    :: ||  |  : ||       | : | :: || |:
Db     329 VPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTT 388

Qy     143 TTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTT------TQAPTTTQAPTTTQAPTT- 196
           :: ||  | : :||  :    | : | :: || |      : || |:    :: || |
Db     389 ESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTS 448

Qy     196 -ITQAATTPAATTPAATTPAATTP--AATTPAATTPGVPAP-------TSAPVWPPICEL 245
            |:::: |  |  ::|| ::: |  ::|| ::: | || |       :||||
Db     449 STTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAP-VPTPSSSTTESSSAPV------- 501

Qy     246 LPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDG 305
                                     ::  |     |             :|::   :
Db     501 ---------------------------TSSTTESSSAP----------VPTPSSSTTES 522

Qy     306 EISPAP-PVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRC 364
           :||| | :   |     :
Db     523 SSAPAPTPSSSTTESSSAPV---------------------------------------- 543

Qy     365 GAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPAD 424
                              |    ||    |                 | |:    :
Db     543 --------------------TSSTTESSSAPV-----------------PTPSSSTTE 563

Qy     425 FSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQ 484
           |                    : |       |  |:|  :||          | :
Db     564 SS----------------------STPVT-----SSTTESSSAPVPT-------PSSST 588

Qy     485 CTSTAAPTAAP-------TAAPTAAPTAAPTAAPSTVVPPATPPATAAPVP-------PT 530
           |::||   |       ::||   |::: | : |  |  :|  :::|||         :
Db     589 TESSSAPVPTPSSSTTESSSAPAPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTES 648

Qy     531 TAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPIT-----VTSAPT 585
           :: | | |::: | : :   | :|:: |   ::||  : ||     |:|      :|||
Db     649 SSAPVPTPSSSTTESSSAPVP-TPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPV 707

Qy     586 AAPTAAPT---AAPTAAPTTAVPE-------IPTTVTSPPTAAPTTAAPAPNTTVTVPPT 635
             |::: |   :||   |::: |         |:: |:  ::|| |::   :::  ||
Db     708 PTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVPTP 767

Qy     636 AAPTT---AAPAPN-TTVTVPPTAAPTAAP------PTVAHAPNTTAAPVTTTSAPATTP 685
           :: ||   :|| |  :: |   ::||   |       :||  |  :::   |:|||::||
Db     768 SSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSVAPVPTPSSSSNITSSAPSSTP 827

Qy     686 -----EDDDIDPPLPND--------PINPCVEECNVLP 710
                |   :  | |:         |::    | :| |
Db     828 FSSSTESSSVPVPTPSSSTTESSSAPVSSSTTESSVAP 865


RESULT  11
S49915
extensin-like protein - maize
C;Species: Zea mays (maize)
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 09-Sep-1997
C;Accession: S49915
R;Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A;Description: Pex genes: pollen-specific genes with extensin-like domains.
A;Reference number: S49915
A;Accession: S49915
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1188 <RUB>
A;Cross-references: EMBL:Z34465; NID:g600117; PID:g600118

  Query Match             9.1%;  Score 392;  DB 2;  Length 1188;
  Best Local Similarity   22.5%;  Pred. No. 6.6e-13;
  Matches  170;  Conservative   87;  Mismatches  324;  Indels  176;  Gaps   28;

Qy       2 IKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDC-----PPAEQHWLLPHEYDCTKFYY 56
           :||      :|  :  | ||     | :  | |       || |:    |
Db     520 VKTTSPPAPIGSPSPPPPVSVVSPPPPVKSPPPPAPVGSPPPPEKSPPPP---------- 570

Qy      57 CEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAP------- 110
                 ||     |       | | :     | |||     |   ::|
Db     570 --------APVASPPPPVKSPPPPTLVASPPPPVKSPPPPAPVASPPPPVKSPPPPTPVA 621

Qy     110 -TTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAP 168
            ||  ::  |  :   ||  :  |   ::|       :|  |      ||:   :|
Db     622 SPPPPAPVASSPPPMKSPPPPTPVSSPPPPEKSPPPPPPAKSTPPPEEYPTPPTSVKSSP 681

Qy     169 TTTTQAPTTTQAP-----------TTTQAPTTTQAPTTITQAATTPAATTPAATTPA-- 215
           |  |  |           | :: |:: : |:   :  ::|  |  ::  ||
Db     682 PPEKSLPPPTLIPSPPPQEKPTPPSTPSKPPSSPEKPSPPKEPVSSPPQTPKSSPPPAPV 741

Qy     215 -------ATTPAATTPAATTPGV-----PAPTSAPVWPPICELLPNGCPADFDIHLLIPH 262
                 ::| |  | :: | |     ||| |:|  ||  ::     |    :    |
Db     742 SSPPPTPVSSPPALAPVSSPPSVKSSPPPAPLSSP--PPAPQV--KSSPPPVQVSSPPPA 797

Qy     263 DKYCNLFYQCSNGYTFEQRC-PEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDED 321
           |         |:    |:   |     :| :    || :|     :|  ||| : :
Db     798 PKSSPPLAPVSSPPQVEKTSPPPAPLSSPPLAPKSSPPHV----VVSSPPPVVKSSPP-- 852

Qy     322 IDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIE 381
                 ||      | |                   :      | |          |
Db     852 -----------PAPVSSPPLTP-----------------KPASPPAHVS------SPPE 876

Qy     382 LVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYL 441
           :|  : |   :  :             |:||    |      |   |:   :   |
Db     877 VVKPSTPPAPTTVISP----------PSEPKSSPP-----PTPVSLPPPIVKSS------ 916

Qy     442 QCVHGQTIARPCPGNLHFSPAT-QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPT 500
                 | |  :   | | :|   ||          |    |  ::|  || ::|
Db     916 ----------PPPAMVSSPPMTPKSSPPPVVV--------SSPPPTVKSSPPPAPVSSPP 957
```

```
Qy    501 AAPTAAPTAAPSTVVPP---ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTV 557
Db    958 ATPKSSPPPAPVNLPPPEVKSSPPPTPVSSPP------PAPKSSPPPAPMSSPP--PPEV 1009

Qy    558 TVPPTAAPTAAPTTAV--PEIPITVTSAP---TAAPTAAPTAAP-----TAAPTTAVPEI 607
Db   1010 KSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPISSP 1069

Qy    608 PTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHA 667
Db   1070 PPPVKSPPPPAPVSSPPPP---VKSPPPPAPVSSPPPP---IKSPPPPAPVSSPPPAPVK 1123

Qy    668 PNTTAAPVTTTS-APATTP-----EDDDIDPPLPNDP 698
Db   1124 PPSLPPPAPVSSPPPVVTPAPPKKEEQSLPPPAESQP 1160
```

RESULT 12
A37232
mucin, tracheal (AMN-22) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 30-Sep-1993
C;Accession: A37232
R;Gerard, C.; Eddy Jr., R.L.; Shows, T.B.
J. Clin. Invest. 86, 1921-1927, 1990
A;Title: The core polypeptide of cystic fibrosis tracheal mucin contains a tandem repeat
A;Reference number: A37232; MUID:91072667
A;Accession: A37232
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-294 <GER>
A;Cross-references: GB:M57417

Query Match 8.8%; Score 381; DB 2; Length 294;
Best Local Similarity 23.8%; Pred. No. 7.3e-13;
Matches 137; Conservative 9; Mismatches 135; Indels 294; Gaps 14;

```
Qy    100 TQAPATTQAPTTTQAPTTTTQ---APT-TTTQAP-----TTTTQAPTTTQAPTTTQAPTT 150
Db      3 TPTPTGTQTPTPTPITTTTTTMVTPTPTIITSTQTPTPTPITTTTVTPTPT--PTSTQRTTP 60

Qy    151 TQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAAT-TPAATTPA 209
Db     61 TSITTTTTVTPTPT---PITTTTTVTPTPTPTGTQTPTTTPISTTTTVTPTPTPTGTQTL 117

Qy    210 ATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLF 269
Db    118 TPTPITTT-TTVTPTPTPTGTQTPTSTPI------------------------------ 146

Qy    270 YQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLD 329
Db    146 ------------------------------------------------------------ 146

Qy    330 NGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPG 389
Db    146 ------------------------------------------------------------ 146

Qy    390 GESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTI 449
Db    146 ------------------------------------------------------------ 146

Qy    450 ARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTA 509
Db    146 ----------------------------------------TTTTTVTPTPTPTGTQT- 163

Qy    510 APSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAP 569
Db    163 ------PTTTPITTTTTVTPT---PTPTGTRYPTPTPIT------TTTTVTPTPTPTSTQ 207

Qy    570 TTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPT-TVTSPPTAAPTTAAPAPNT 628
Db    208 STTPTPITTTNTVTPTPTPTGTQTPTPTPITTTTTMVTPTPTITSTQTPTPT-----PIT 262
```

```
Qy    629 TVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPT 663
Db    263 TTTVTPTPTPTSTQRTTPTSITTTTTVTPT---PT 294
```

RESULT 13
A48292
mucin, tracheobronchial - dog
C;Species: Canis lupus familiaris (dog)
C;Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 10-Sep-1997
C;Accession: A48292
R;Verma, M.; Davidson, E.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 7144-7148, 1993
A;Title: Molecular cloning and sequencing of a canine tracheobronchial mucin cDNA con
A;Reference number: A48292
A;Accession: A48292
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1118 <VER>
A;Cross-references: GB:X69164; NID:g402557; PID:g402558
A;Note: authors translated the codon CAA for residue 52 as Gly, GAT for residue 94 as
C;Comment: It is uncertain whether Met-1 or Met-52 is the initiator.
C;Keywords: glycoprotein

Query Match 8.4%; Score 361.5; DB 2; Length 1118;
Best Local Similarity 23.3%; Pred. No. 2.1e-11;
Matches 207; Conservative 49; Mismatches 295; Indels 339; Gaps 37;

```
Qy     66 PRDCAPGTEFKFSAQTCVHAALAGCTLP----------GPPAETTQAPATTQAPTTTQAP 115
Db    203 PPQCLTPTPTPLSTPSIIRPQPEALTLPLCLSAVSETTPTPISTTTTVTPTPTPTGTQTP 262

Qy    116 TTTTQAPTTTTQAPTTT-TQAPTT---------------------------------TQ 140
Db    263 TTTPITTTTVTPTPTPTGTQTPTTVLITTTTTMPTPTPTSQKLRGSDTPVLVIALVTPTP 322

Qy    141 APTTTQAPTTTQAPTTTTQAPTTT---TQAPTTTTQAPTTTQAPT--------------- 183
Db    323 TPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTIRITTTTTTMTPTPTTTSTKVLVTPYTT 382

Qy    183 ------TTQAPTTTQAPTTITQAATTPAATTPAATTPA----------------ATTPAA 220
Db    383 TTGVTPTTHRHTRHQTTTTHTTYTTTVTAIATPTGTRPTTYLHTTTTVTQTTHRQTTRPQP 442

Qy    221 TTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQ 280
Db    443 TTPSPTTTTVD-PTQHPSPPPLGDPNPN-THRDTDPN----HDTTTTT--TVTPTPTPTR 494

Qy    281 RCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNED----------------EDI 322
Db    495 TRPQPRYYTTTTTMTPTP---------TPPTPHTDHHHDPLPPPPPQVRHHRTPRHCHEL 545

Qy    323 DIGDLL---DNGCPANFEIDWLLPHGNRCDKYYQCVH------------GNLVERRCGAG 367
Db    546 TSGSPLTHYPNTPTTN------LLEKNRSELINTPTHTMIISSAAMATPPSLVHSRCRSG 599

Qy    368 THFSFELQQ-----CDHIEL-------VGCTLPGG-------------ESEEVDVDEDAC 402
Db    600 -HSSKAQRKSRASGCGLMMLGVESGVGVGSTHAGGALVRSRCRFTRHSRAEE-------- 651

Qy    403 TGWYCPTEPIEWEPLPNGCPADFSIDHLLPHES-------DCGQYLQCVHGQTIARPCPG 455
Db    651 ----------------GLGPEAVVSMMLGVESGVGVGVRHCGGRLHSAGGG--AEDDPG 691

Qy    456 NL-----------HFSPATQSCESPV-----------TAGCQVFECDSDNQCTSTAAPTA 493
Db    692 RVPRVGALTDSTQHTSRKAIKCKAGVPNEELTTYIACASALQSETCAADMIATATIHTTT 751

Qy    494 APT-------AAPTAAPTAAPTAAPSTVV----------------PPATPPATAAPVPPT 530
Db    752 SRSIRKPKTISELTTYACASTAASVGTVVQLDMIATGGLRDRGQMAGAMPAGDSRGSPGT 811
```

```
Qy    531 TAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTA 590
Db    812 ELEFRPNGLQAPTPTPIS------TTTTVTPTPTPTG------------TQTPTTTPSP 852

Qy    591 APTAAPTAAPTT--AVPEIPTTVTSPPTAAPTT---------------------------- 622
Db    853 PPPGDPTKGKTVYKIIGSLRTTFNLPAWLGYITICKCLMQSNRKTKTAETINVKTLGSLM 912

Qy    622 --AAPAPNTTVTVP--------------PTA----------------APT-------TAA 642
Db    913 MDVTPINSCRVTPPFPSGKPCADSRLMNPTAETGSVIGAGCSCCVLTTPTTHRVIRCTTA 972

Qy    643 PAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDP 692
Db    973 PTP-TPISTTTTVTPTQHPP--AHRP--TTSPITTTTT-VTQPDTTHTDP 1016


RESULT 14
A38420
antifreeze glycoprotein precursor - black rockcod
C;Species: Notothenia coriiceps neglecta (black rockcod)
C;Date: 28-Jun-1991 #sequence_revision 30-Jan-1993 #text_change 09-Sep-1997
C;Accession: A38420
R;Hsiao, K.C.; Cheng, C.H.C.; Fernandes, I.E.; Detrich, H.W.; DeVries, A.L.
Proc. Natl. Acad. Sci. U.S.A. 87, 9265-9269, 1990
A;Title: An antifreeze glycopeptide gene from the antarctic cod Notothenia coriiceps neg
A;Reference number: A38420; MUID:91067687
A;Accession: A38420
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-822 <HSI>
A;Cross-references: GB:M55000; NID:g213391; PID:g213392
A;Note: the authors translated the codon CCT for residue 489 as Ala
C;Keywords: glycoprotein


  Query Match             8.2%;  Score 354;  DB 2;  Length 822;
  Best Local Similarity   25.2%;  Pred. No. 3.8e-11;
  Matches  185;  Conservative   21;  Mismatches  319;  Indels  210;  Gaps   25;

Qy     65 APRDCAPGTEFKFSAQTCVHAA---------LAGCTLPGPP----AETTQAPATTQAPTT 111
Db    146 AATPATPAFHFAATAATPATAATPALIFAATAATAATPATPAFHFAATAATPATAATPAL 205

Qy    112 TQAPTTTTQA-PTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTQAPTT 170
Db    206 IFAATAATAATPATAALNFAATAATPATAATPALIFAATAATAATPATAALNFAATAATA 265

Qy    171 TTQAPTTTQAPTTTQAPTTTQAPTTITQA----ATTPA---------ATTPA-ATTP--- 214
Db    266 ATPATAACNFAATAATPATAATPALIFAATAATAATPATAACNFAATAATPATAATPALI 325

Qy    214 -AATTPAATTPA------ATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYC 266
Db    326 FAATAATAATPATAALNFAATAATPATAATP---------------------------- 357

Qy    267 NLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGD 326
Db    357 -----------------ALIFAATAATAATPATAALNFAATAATPATAAT-PALIFAAT 397

Qy    327 LLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGT----HFSFELQQCDHIEL 382
Db    398 AATAATPATAAL------------------NFAATAATAATPATPAFNF--------AA 430

Qy    383 VGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQ 442
Db    431 TAATPATAATPALIFAATAATA-----------ATPATAALNFAATAATPATAATPALIF 479

Qy    443 CVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAA 502
Db    480 AATAATAATPATAALHFAATAATAATPATAAL--------NFAATAATPATAATPALIFA 531

Qy    503 PTAAPTAAPST-------VVPPATPPATA-------APVPPTTAIPTP--APTAAPTAAP 546
Db    532 ATAATAATPATAAFNFAATAATAATPATAALNFAATAATPATAATPALIFAATAATAATP 591

Qy    547 TTAAPESPTTVTVPPTAAPTA------APTTAVPEIP---------------------- 578
Db    592 ATAALNFAATAATPATAATPALIFAATAATAATPATPAFHFAATAATPATAATPALIFAA 651

Qy    578 --ITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAA-PTTAAPAPNTTVTVPP 634
Db    652 TAATAATPATAALNFAATAATAATPATAALNFAATAATPATAATPALIFAATAATAATPA 711

Qy    635 TAA---------PTTAA------PAPNTTVTVPPTA-------APTAAPPTVAHAP---- 669
Db    712 TAALNFAATAATPATAATPALIFAATAATAATPATAAFNFAATAATAATPATAATPALIF 771

Qy    669 NTTAAPVTTTSAPAT 683
Db    772 AATAATAATPATPAT 786


RESULT 15
S15921
protein TPX-VT3 - Thermoproteus phage 1 (strain strain KRA1 10/12)
C;Species: Thermoproteus phage 1, TTV1
A;Variety: strain KRA1 10/12
C;Date: 21-Nov-1993 #sequence_revision 26-May-1995 #text_change 09-Sep-1997
C;Accession: S15921; S06686
R;Neumann, H.; Zillig, W.
Nucleic Acids Res. 18, 2171, 1990
A;Title: Nucleotide sequence of the viral protein TPX of the TTV1 variant VT3.
A;Reference number: S15921; MUID:90245666
A;Accession: S15921
A;Molecule type: DNA
A;Residues: 1-474 <NEU>
A;Cross-references: EMBL:X14717; NID:g62191; PID:e20711; PID:g584349
A;Experimental source: strain KRA1 10/12
A;Note: the authors translated the initiation codon GTG for residue 1 as Val
R;Neumann, H.
submitted to the EMBL Data Library, March 1989
A;Reference number: S06686
A;Accession: S06686
A;Molecule type: DNA
A;Residues: 207-474 <NEW>
A;Cross-references: EMBL:X14717
A;Experimental source: strain KRA1 10/12
C;Genetics:
A;Start codon: GTG


  Query Match             8.1%;  Score 351;  DB 2;  Length 474;
  Best Local Similarity   32.6%;  Pred. No. 3.4e-11;
  Matches  103;  Conservative   20;  Mismatches  129;  Indels   64;  Gaps   13;

Qy    394 EVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPC 453
Db    192 EIDISQFAAF-----TDPLVMYLSINGSTANVKV-------YKQGSNIGTVSGNYSTTPY 239

Qy    454 PGN-----------------LHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAP 495
Db    240 -GNPSMAGYGTVDKHYANFIVLPYEPDPQVTVTPI----------SSPSPTPTPTPTPTP 288

Qy    496 TAAPTAAPTAAPTAAPS---TVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPE 552
Db    289 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTP 348

Qy    553 SPT-TVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTTAVPEIPT-T 610
Db    349 TPTPTPTPTPTPTPTPTPTYDITYVVFDVTPSPTPTPTPTPTPTPTPT------PTPT 402

Qy    611 VTSPPTAAPTTAAPAPNTTVTVPPTAAPT-TAAPAPNTTVT-----VPPTAAPTAAPPTV 664
Db    403 PTPTPTPTPT---PTPTPTPTPTPTPTPTPTPTPTYDITYVIFDVTPSPTPT---PTP 456
```

```
Qy     665 AHAPNTTAAPVTTTSA 680
               |  |  | :|||:
Db     457 TPTPTPTPTPTSTTSS 472
```

Search completed: July 18, 1999, 00:45:05
Job time: 261 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 03:21:16 ; Search time 42.46 Seconds
(without alignments)
901.494 Million cell updates/sec

Title: US-09-294-663-3
Perfect score: 4328
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 788

Scoring table: BLOSUM62

Searched: 396714 seqs, 48575446 residues

Database : Pending_Patents_AA:*
1: /cgn2_6/ptodata/1/paa/PCTUS9_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US084A_COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep:*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/1/paa/US60_COMB.pep:*
19: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep:*
20: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*
21: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*
22: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*
23: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4328	100.0	788	16	US-09-294-663-3	Sequence 3, Appli
2	4293	99.2	786	15	US-09-103-429A-3	Sequence 3, Appli
3	4266.5	98.6	807	16	US-09-294-663-4	Sequence 4, Appli
4	4231.5	97.8	805	15	US-09-103-429A-4	Sequence 4, Appli
5	521.5	12.0	1205	2	US-07-643-502C-28	Sequence 28, Appl
6	520.5	12.0	1404	2	US-07-643-502C-13	Sequence 13, Appl
7	520.5	12.0	1404	2	US-07-757-022B-2	Sequence 2, Appli
8	520.5	12.0	941	2	US-07-757-022B-14	Sequence 14, Appl
9	520.5	12.0	1361	2	US-07-757-022B-40	Sequence 40, Appl
10	520.5	12.0	1311	2	US-07-757-022B-42	Sequence 42, Appl
11	520.5	12.0	1270	2	US-07-757-022B-44	Sequence 44, Appl
12	520.5	12.0	1320	2	US-07-757-022B-46	Sequence 46, Appl
13	520.5	12.0	1354	2	US-07-757-022B-48	Sequence 48, Appl
14	520.5	12.0	1314	2	US-07-757-022B-50	Sequence 50, Appl
15	520.5	12.0	1363	2	US-07-757-022B-52	Sequence 52, Appl
16	520.5	12.0	1049	2	US-07-757-022B-58	Sequence 58, Appl
17	520.5	12.0	1320	2	US-07-757-022B-60	Sequence 60, Appl
18	520.5	12.0	1404	2	US-07-757-022B-62	Sequence 62, Appl
19	520.5	12.0	1038	2	US-07-757-022B-74	Sequence 74, Appl
20	520.5	12.0	1022	2	US-07-757-022B-84	Sequence 84, Appl
21	520.5	12.0	1140	2	US-07-757-022B-104	Sequence 104, App
22	520.5	12.0	1313	2	US-07-757-022B-142	Sequence 142, App
23	455.5	10.5	1837	13	US-08-928-361-5	Sequence 5, Appli
24	455.5	10.5	1837	21	US-08-928-361B-5	Sequence 5, Appli
25	445.5	10.3	1721	1	PCT-US97-14104-5	Sequence 5, Appli
26	445.5	10.3	1721	11	US-08-700-651A-5	Sequence 5, Appli
27	445.5	10.3	1721	11	US-08-700-651-5	Sequence 5, Appli
28	445.5	10.3	1721	13	US-08-928-361-6	Sequence 6, Appli
29	445.5	10.3	1721	21	US-08-928-361B-6	Sequence 6, Appli
30	393.5	9.1	1367	18	US-60-087-236-55	Sequence 55, Appl
31	373.5	8.6	557	16	US-09-248-796-26892	Sequence 26892, A
32	373.5	8.6	557	18	US-60-096-409-26892	Sequence 26892, A
33	333.5	7.7	827	16	US-09-248-796-17307	Sequence 17307, A
34	333.5	7.7	827	18	US-60-096-409-17307	Sequence 17307, A
35	322.5	7.5	3880	3	US-08-028-021-1	Sequence 1, Appli
36	317	7.3	297	9	US-08-580-545B-6	Sequence 6, Appli
37	317	7.3	297	16	US-09-262-653A-6	Sequence 6, Appli
38	313	7.2	907	11	US-08-783-774-2	Sequence 2, Appli
39	309.5	7.2	216	13	US-08-928-361-8	Sequence 8, Appli
40	309.5	7.2	216	21	US-08-928-361B-8	Sequence 8, Appli
41	301	7.0	216	21	US-08-928-361B-27	Sequence 27, Appl
42	298.5	6.9	288	16	US-09-216-393-341	Sequence 341, App
43	298.5	6.9	288	16	US-09-216-393-344	Sequence 344, App
44	296	6.8	249	11	US-08-700-651-15	Sequence 15, Appl
45	296	6.8	249	13	US-08-928-361-20	Sequence 20, Appl

ALIGNMENTS

```
RESULT 1
US-09-294-663-3
; Sequence 3, Application US/09294663
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga Street
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/294,663
;      FILING DATE:  19-APR-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/103,429
;      FILING DATE:  24-JUN-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39-CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  788 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
```

```
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  peritrophic membrane
US-09-294-663-3


  Query Match             100.0%;  Score 4328;  DB 16;  Length 788;
  Best Local Similarity   100.0%;  Pred. No. 1e-271;
  Matches  788;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60

Qy     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180

Qy    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240

Qy    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300

Qy    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360

Qy    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420

Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAA 660

Qy    661 PPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHADCDKYW 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 PPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHADCDKYW 720

Qy    721 VCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMSESYEGVQVFIPWNKLDEDI 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 VCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMSESYEGVQVFIPWNKLDEDI 780

Qy    781 RQALNFEL 788
          ||||||||
Db    781 RQALNFEL 788


RESULT   2
US-09-103-429A-3
; Sequence 3, Application US/09103429A
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,429A
;      FILING DATE:  24-JUN-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  786 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  peritrophic membrane
US-09-103-429A-3


  Query Match             99.2%;  Score 4293;  DB 15;  Length 786;
  Best Local Similarity   99.7%;  Pred. No. 1.9e-269;
  Matches  786;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60

Qy     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180
          |||||||| |||||||||||||||| |||||||||||||||||||||||||||||
Db    121 APTTTTQA-TTTTQAPTTTQAPTTTQA-TTTQAPTTTTQAPTTTTQAPTTTTQA 178

Qy    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 238

Qy    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 298

Qy    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 358

Qy    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 418
```

```
Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 478

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    479 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 538

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 598

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    599 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAA 658

Qy    661 PPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHADCDKYW 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    659 PPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHADCDKYW 718

Qy    721 VCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMSESYEGVQVFIPWNKLDEDI 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    719 VCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMSESYEGVQVFIPWNKLDEDI 778

Qy    781 RQALNFEL 788
          ||||||||
Db    779 RQALNFEL 786
```

```
RESULT   3
US-09-294-663-4
; Sequence 4, Application US/09294663
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga Street
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/294,663
;      FILING DATE:  19-APR-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/103,429
;      FILING DATE:  24-JUN-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39-CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  807 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  peritrophic membrane
US-09-294-663-4


  Query Match             98.6%;  Score 4266.5;  DB 16;  Length 807;
  Best Local Similarity   97.1%;  Pred. No. 9.8e-268;
  Matches  784;  Conservative    0;  Mismatches    4;  Indels   19;  Gaps    1;

Qy      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60
          |||||||||||||||||||||||||||||||||||||| |||||  ||| ||||||||||||
Db      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYG 60

Qy     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180

Qy    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240

Qy    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300

Qy    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360

Qy    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420

Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTV----------------PPTAAPTTA 641
          |||||||||||||||||||||||||||||||                 |||||||||
Db    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 660

Qy    642 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 701
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 720

Qy    702 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 780

Qy    762 ESYEGVQVFIPWNKLDEDIRQALNFEL 788
          |||||||||||||||||||||||||||
Db    781 ESYEGVQVFIPWNKLDEDIRQALNFEL 807
```

```
RESULT   4
US-09-103-429A-4
; Sequence 4, Application US/09103429A
;. GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
```

```
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,429A
;      FILING DATE:  24-JUN-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  805 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  peritrophic membrane
US-09-103-429A-4


  Query Match             97.8%;  Score 4231.5;  DB 15;  Length 805;
  Best Local Similarity   96.9%;  Pred. No. 1.8e-265;
  Matches  782;  Conservative    0;  Mismatches    4;  Indels   21;  Gaps    3;

Qy      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60
          ||||||||||||||||||||||||||||||| ||||| ||| ||||||||||||
Db      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYG 60

Qy     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQA 180
          |||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    121 APTTTTQA-TTTTQAPTTTQAPTTTQA-TTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 178

Qy    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 238

Qy    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 298

Qy    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 358

Qy    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 418

Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 478

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    479 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 538

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 598

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTV----------------------PPTAAPTTA 641
          |||||||||||||||||||||||||||||||                      |||||||||
Db    599 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 658

Qy    642 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 701
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    659 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 718

Qy    702 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    719 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 778

Qy    762 ESYEGVQVFIPWNKLDEDIRQALNFEL 788
          |||||||||||||||||||||||||||
Db    779 ESYEGVQVFIPWNKLDEDIRQALNFEL 805


RESULT   5
US-07-643-502C-28
; Sequence 28, Application US/07643502C
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  A Megakaryocytopoietic Factor
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/643,502C
;      FILING DATE:  19910118
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5166D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  28:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1205 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-07-643-502C-28


  Query Match             12.0%;  Score 521.5;  DB 2;  Length 1205;
  Best Local Similarity   29.5%;  Pred. No. 5.3e-26;
  Matches  203;  Conservative   26;  Mismatches  205;  Indels  253;  Gaps   32;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
           | | || ||: || |      ||| : | :||          ||||  : |||||  ||
Db     132 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 191

Qy     137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
           || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db     192 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 251

Qy     186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
            |||| : ||  |     |    || |||    | |   || |||  ||||: |  |
Db     252 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 309

Qy     244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                    |   :  ||   :
Db     309 ----------------------------------------------PTTTKEPSPTTPK- 322

Qy     304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
               |||  |
Db     322 ----EPAPTTT---------------------------------------------------- 329

Qy     364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                  :|      | |  |    || | |  | .   |
Db     329 --------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 361

Qy     423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
           |  :     | |                    :| |    :| |
Db     362 APTT-----PKEP----------------------APTTPKKPAPTT----------- 382

Qy     483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                |||     |||     ||||   |:   |  | | |   : |||     ||
Db     382 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 438

Qy     536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
             ||||     ||||    :|||   | ||    ||| |  |:  |    |    |||
Db     439 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 496

Qy     590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT------------VTVPP 634
            |||     ||||     |||:  |    |   ||   || ||                | |
Db     497 PAPTTPKETAPTTPKETAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 556

Qy     635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH--------------APNTTAAPVTTTS 679
            ||||   ||| |     ||     || |:                ||  ||  | :|||
Db     557 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 616

Qy     680 ---APATTPEDDDIDP--PLPNDPINP 701
              || |  |     |  | |  |  |
Db     617 DKPAPTTPKETAPTTPKEPAPTTPKKP 643


RESULT   6
US-07-643-502C-13
; Sequence 13, Application US/07643502C
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  A Megakaryocytopoietic Factor
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/643,502C
;      FILING DATE:  19910118
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5166D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1404 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-643-502C-13


  Query Match             12.0%;  Score 520.5;  DB 2;  Length 1404;
  Best Local Similarity   29.5%;  Pred. No. 7.1e-26;
  Matches  203;  Conservative   26;  Mismatches  205;  Indels  253;  Gaps   32;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
           | | || ||: || |      ||| : | :||          ||||  : |||||  ||
Db     331 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 390

Qy     137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
           || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db     391 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 450

Qy     186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
            |||| : ||  |     |    || |||    | |   || |||  ||||: |  |
Db     451 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 508

Qy     244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                    |   :  ||   :
Db     508 ----------------------------------------------PTTTKEPSPTTPK- 521

Qy     304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
               |||  |
Db     521 ----EPAPTTT---------------------------------------------------- 528

Qy     364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                  :|      | |  |    || | |  | .   |
Db     528 --------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 560

Qy     423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
           |  :     | |                    :| |    :| |
Db     561 APTT-----PKEP----------------------APTTPKKPAPTT----------- 581

Qy     483 NQCTSTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                |||     |||     ||||   |:   |  | | |   : |||     ||
```

```
Db    581 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 637

Qy    536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
          ||||     ||||    :|||   | ||    ||| |  |: |   |    |||
Db    638 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 695

Qy    590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT-----------VTVPP 634
          |||     ||||     |||:  |    |   ||   || ||              | |
Db    696 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 755

Qy    635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH-------------APNTTAAPVTTTS 679
            ||||   ||| |     ||     || |:                || ||  | :|||
Db    756 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 815

Qy    680 ---APATTPEDDDIDP--PLPNDPINP 701
             || |  |     |  | |  |  |
Db    816 DKPAPTTPKETAPTTPKEPAPTTPKKP 842


RESULT   7
US-07-757-022B-2
; Sequence 2, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1404 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-2

  Query Match            12.0%;  Score 520.5;  DB 2;  Length 1404;
  Best Local Similarity   29.5%;  Pred. No. 7.1e-26;
  Matches  203;  Conservative   26;  Mismatches  205;  Indels  253;  Gaps   32;

Qy     93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
          | | || ||: || |      ||| : | :||           ||||  : |||||  ||
Db    331 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 390

Qy    137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
          || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db    391 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 450

Qy    186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
           |||| : ||  |     |    || |||    | |   || |||  ||||: |  |
Db    451 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 508

Qy    244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                    |   :  ||   :
Db    508 ----------------------------------------------PTTTKEPSPTTPK- 521

Qy    304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
          |||  |
Db    521 ----EPAPTTT---------------------------------------------------- 528

Qy    364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                    :|      | |  |    || | |  |     |
Db    528 -------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 560

Qy    423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
          |  :    | |                            :| |    :| |
Db    561 APTT-----PKEP-----------------------APTTPKKPAPTT----------- 581

Qy    483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                |||     |||     ||||   |:   |  | | |   : |||     ||
Db    581 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 637

Qy    536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
            ||||     ||||    :|||   | ||    ||| |  |: |   |    |||
Db    638 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 695

Qy    590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT-----------VTVPP 634
          |||     ||||     |||:  |    |   ||   || ||              | |
Db    696 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 755

Qy    635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH-------------APNTTAAPVTTTS 679
            ||||   ||| |     ||     || |:                || ||  | :|||
Db    756 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 815

Qy    680 ---APATTPEDDDIDP--PLPNDPINP 701
             || |  |     |  | |  |  |
Db    816 DKPAPTTPKETAPTTPKEPAPTTPKKP 842


RESULT   8
US-07-757-022B-14
; Sequence 14, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  941 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-14


  Query Match           12.0%;  Score 520.5;  DB 2;  Length 941;
  Best Local Similarity   29.5%;  Pred. No. 4.8e-26;
  Matches  203;  Conservative   26;  Mismatches  205;  Indels  253;  Gaps   32;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
           | | || ||: || |      ||| : | :||           ||||  : |||||  ||
Db     132 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 191

Qy     137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
           || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db     192 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 251

Qy     186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
            |||| :  ||  |     |    || |||    | |   || |||  ||||: |  |
Db     252 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 309

Qy     244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                    |   :  ||   :
Db     309 ----------------------------------------------PTTTKEPSPTTPK- 322

Qy     304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
               |||  |
Db     322 ----EPAPTTT---------------------------------------------------- 329

Qy     364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                   :|      | |  |    || | |  |     |
Db     329 ---------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 361

Qy     423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
           |  :     | |                    :| |    :| |
Db     362 APTT-----PKEP-----------------------APTTPKKPAPTT----------- 382

Qy     483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                ||| |||     ||||   |:   |  | | |   : |||     ||
Db     382 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 438

Qy     536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
             ||||     ||||    :|||   | ||    ||| |  |: |   |    |||
Db     439 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 496

Qy     590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT-----------VTVPP 634
              |||     ||||     |||:  |    |   ||   || ||             | |
Db     497 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 556

Qy     635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH-------------APNTTAAPVTTTS 679
              ||||   ||| |     ||     || |:                || ||  | :|||
Db     557 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 616

Qy     680 ---APATTPEDDDIDP--PLPNDPINP 701
               || |  |     |  | |  |  |
Db     617 DKPAPTTPKETAPTTPKEPAPTTPKKP 643


RESULT   9
US-07-757-022B-40
; Sequence 40, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1361 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-40


  Query Match           12.0%;  Score 520.5;  DB 2;  Length 1361;
  Best Local Similarity   29.5%;  Pred. No. 6.9e-26;
  Matches  203;  Conservative   26;  Mismatches  205;  Indels  253;  Gaps   32;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
```

```
          | | || ||: || |      ||| : | :||          ||||  : |||||  ||
Db    288 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 347

Qy    137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
          || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db    348 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 407

Qy    186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
           |||| : ||  |     |    || |||    | |   || |||  ||||: |  |
Db    408 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 465

Qy    244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                          |   :  ||   :
Db    465 ----------------------------------------------PTTTKEPSPTTPK- 478

Qy    304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
               |||  |
Db    478 ----EPAPTTT---------------------------------------------------- 485

Qy    364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                    :|      | |  |    || | |  |     |
Db    485 --------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 517

Qy    423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
          |  :     | |                              :| |    :| |
Db    518 APTT-----PKEP-----------------------APTTPKKPAPTT----------- 538

Qy    483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                |||     |||     ||||   |:   |  | | |   : |||     ||
Db    538 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 594

Qy    536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
            ||||     ||||    :|||   | ||    ||| |  |: |   |    |||
Db    595 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 652

Qy    590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT------------VTVPP 634
           |||     ||||     |||:  |    |   ||   || ||                | |
Db    653 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 712

Qy    635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH--------------APNTTAAPVTTTS 679
           ||||   ||| |     ||     || |:                 || ||  | :|||
Db    713 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 772

Qy    680 ---APATTPEDDDIDP--PLPNDPINP 701
             || |  |     |  | |  |  |
Db    773 DKPAPTTPKETAPTTPKEPAPTTPKKP 799


RESULT  10
US-07-757-022B-42
; Sequence 42, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1311 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-42


  Query Match             12.0%;  Score 520.5;  DB 2;  Length 1311;
  Best Local Similarity   29.5%;  Pred. No. 6.7e-26;
  Matches  203;  Conservative   26;  Mismatches  205;  Indels  253;  Gaps   32;

Qy     93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
          | | || ||: || |      ||| : | :||          ||||  : |||||  ||
Db    238 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 297

Qy    137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
          || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db    298 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 357

Qy    186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
           |||| : ||  |     |    || |||    | |   || |||  ||||: |  |
Db    358 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 415

Qy    244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                          |   :  ||   :
Db    415 ----------------------------------------------PTTTKEPSPTTPK- 428

Qy    304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
               |||  |
Db    428 ----EPAPTTT---------------------------------------------------- 435

Qy    364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                    :|      | |  |    || | |  |     |
Db    435 --------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 467

Qy    423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
          |  :     | |                              :| |    :| |
Db    468 APTT-----PKEP-----------------------APTTPKKPAPTT----------- 488

Qy    483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                |||     |||     ||||   |:   |  | | |   : |||     ||
Db    488 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 544

Qy    536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
            ||||     ||||    :|||   | ||    ||| |  |: |   |    |||
Db    545 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 602

Qy    590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT------------VTVPP 634
           |||     ||||     |||:  |    |   ||   || ||                | |
Db    603 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 662
```

```
Qy    635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH-------------APNTTAAPVTTTS 679
          ||||   ||| |      ||     || |:               || ||  | :|||
Db    663 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 722

Qy    680 ---APATTPEDDDIDP--PLPNDPINP 701
            || |  |     |  | |  |  |
Db    723 DKPAPTTPKETAPTTPKEPAPTTPKKP 749


RESULT  11
US-07-757-022B-44
; Sequence 44, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1270 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-44


  Query Match          12.0%;  Score 520.5;  DB 2;  Length 1270;
  Best Local Similarity   29.5%;  Pred. No. 6.5e-26;
  Matches  203;  Conservative   26;  Mismatches  205;  Indels  253;  Gaps   32;

Qy     93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
          | | || ||: || |      ||| : | :||          ||||  : ||||| ||
Db    197 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 256

Qy    137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
          || :  |||||:  |||| :  ||||| : ||||| ||||  : |||| :  |||| :

Db    257 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 316

Qy    186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
           |||| : ||  |      |    || |||    | |   || |||  ||||: | |
Db    317 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 374

Qy    244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                        |   :  ||   :
Db    374 ----------------------------------------------PTTTKEPSPTTPK- 387

Qy    304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
              |||  |
Db    387 ----EPAPTTT---------------------------------------------------- 394

Qy    364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                      :|      | |  |    || | |  |     |
Db    394 --------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 426

Qy    423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
          |  :     | |                              :| |    :| |
Db    427 APTT-----PKEP-----------------------APTTPKKPAPTT----------- 447

Qy    483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                ||      |||     ||||   |:   |  | | |   : |||     ||
Db    447 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 503

Qy    536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
            ||||     ||||    :|||   | ||    ||| |  |: |   |    |||
Db    504 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 561

Qy    590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT------------VTVPP 634
           |||     ||||     |||:  |    |   ||   || ||              | |
Db    562 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 621

Qy    635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH-------------APNTTAAPVTTTS 679
            ||||   ||| |      ||     || |:               || ||  | :|||
Db    622 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 681

Qy    680 ---APATTPEDDDIDP--PLPNDPINP 701
            || |  |     |  | |  |  |
Db    682 DKPAPTTPKETAPTTPKEPAPTTPKKP 708


RESULT  12
US-07-757-022B-46
; Sequence 46, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1320 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-46


  Query Match             12.0%;  Score 520.5;  DB 2;  Length 1320;
  Best Local Similarity   29.5%;  Pred. No. 6.7e-26;
  Matches  203;  Conservative   26;  Mismatches  205;  Indels  253;  Gaps   32;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
           | | || ||: || |      ||| : | :||           ||||  : |||||  ||
Db     247 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 306

Qy     137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
           || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db     307 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 366

Qy     186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
            |||| : ||  |     |    || |||    | |   || |||  ||||: |  |
Db     367 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 424

Qy     244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                    |   :  ||   :
Db     424 ----------------------------------------------PTTTKEPSPTTPK- 437

Qy     304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
                ||| |
Db     437 ----EPAPTTT---------------------------------------------------- 444

Qy     364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                  :|      | |  |     || | |  |      |
Db     444 --------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 476

Qy     423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
           |  :     | |                          :| |    :| |
Db     477 APTT-----PKEP-----------------------APTTPKKPAPTT----------- 497

Qy     483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                ||| |||     ||||   |:   |  | | |   : |||      ||
Db     497 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 553

Qy     536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
             ||||     ||||    :|||   | ||    ||| |  |: |    |     |||
Db     554 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 611

Qy     590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT-----------VTVPP 634
            |||     ||||     |||:  |    |   ||   || ||              | |
Db     612 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 671

Qy     635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH-------------APNTTAAPVTTTS 679
             ||||   ||| |     ||     || |:                || ||  | :|||
Db     672 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 731

Qy     680 ---APATTPEDDDIDP--PLPNDPINP 701
```

```
           || |  |     |  | |  |  |
Db     732 DKPAPTTPKETAPTTPKEPAPTTPKKP 758


RESULT  13
US-07-757-022B-48
; Sequence 48, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1354 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-48


  Query Match             12.0%;  Score 520.5;  DB 2;  Length 1354;
  Best Local Similarity   29.5%;  Pred. No. 6.9e-26;
  Matches  203;  Conservative   26;  Mismatches  205;  Indels  253;  Gaps   32;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
           | | || ||: || |      ||| : | :||           ||||  : |||||  ||
Db     281 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 340

Qy     137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
           || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db     341 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 400

Qy     186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
            |||| : ||  |     |    || |||    | |   || |||  ||||: |  |
Db     401 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 458
```

```
Qy    244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                 |   :  ||   :
Db    458 ----------------------------------------------PTTTKEPSPTTPK- 471

Qy    304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
              |||  |
Db    471 ----EPAPTTT---------------------------------------------------- 478

Qy    364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                      :|      | |  |    || | |  |     |
Db    478 --------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 510

Qy    423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
          |  :     | |                      :| |    :| |
Db    511 APTT-----PKEP----------------------APTTPKKPAPTT----------- 531

Qy    483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                |||     |||     ||||   |:   |  | | |   : |||     ||
Db    531 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 587

Qy    536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
            ||||     ||||    :|||   | ||    ||| |  |: |   |    |||
Db    588 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 645

Qy    590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT-----------VTVPP 634
           |||     ||||     |||:  |    |   ||   || ||             | |
Db    646 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 705

Qy    635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH-------------APNTTAAPVTTTS 679
            ||||   ||| |     ||     || |:                || ||  | :|||
Db    706 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 765

Qy    680 ---APATTPEDDDIDP--PLPNDPINP 701
             || |  |     |  | |  |  |
Db    766 DKPAPTTPKETAPTTPKEPAPTTPKKP 792
```

RESULT 14
US-07-757-022B-50
; Sequence 50, Application US/07757022B
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1314 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-50

Query Match 12.0%; Score 520.5; DB 2; Length 1314;
Best Local Similarity 29.5%; Pred. No. 6.7e-26;
Matches 203; Conservative 26; Mismatches 205; Indels 253; Gaps 32;

```
Qy     93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
          | | || ||: || |       ||| : | :||           ||||  : |||||  ||
Db    241 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 300

Qy    137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
          || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db    301 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 360

Qy    186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
           |||| : ||  |      |    || |||    | |   || |||  ||||: |  |
Db    361 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 418

Qy    244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                 |   :  ||   :
Db    418 ----------------------------------------------PTTTKEPSPTTPK- 431

Qy    304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
              |||  |
Db    431 ----EPAPTTT---------------------------------------------------- 438

Qy    364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                      :|      | |  |    || | |  |     |
Db    438 --------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 470

Qy    423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
          |  :     | |                      :| |    :| |
Db    471 APTT-----PKEP----------------------APTTPKKPAPTT----------- 491

Qy    483 NQCTSTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                |||     |||     ||||   |:   |  | | |   : |||     ||
Db    491 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 547

Qy    536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
            ||||     ||||    :|||   | ||    ||| |  |: |   |    |||
Db    548 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 605

Qy    590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT-----------VTVPP 634
           |||     ||||     |||:  |    |   ||   || ||             | |
Db    606 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 665

Qy    635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH-------------APNTTAAPVTTTS 679
            ||||   ||| |     ||     || |:                || ||  | :|||
Db    666 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 725

Qy    680 ---APATTPEDDDIDP--PLPNDPINP 701
             || |  |     |  | |  |  |
Db    726 DKPAPTTPKETAPTTPKEPAPTTPKKP 752
```

RESULT 15

```
US-07-757-022B-52
; Sequence 52, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1363 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-52


  Query Match            12.0%;  Score 520.5;  DB 2;  Length 1363;
  Best Local Similarity   29.5%;  Pred. No. 6.9e-26;
  Matches  203;  Conservative   26;  Mismatches  205;  Indels  253;  Gaps   32;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
           | | || ||: || |      ||| : | :||          ||||  : |||||  ||
Db     290 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 349

Qy     137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
           || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db     350 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 409

Qy     186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
            |||| : ||  |      |    || |||    | |   || |||  ||||: |  |
Db     410 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 467

Qy     244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                 |   :  ||   :
Db     467 ---------------------------------------------PTTTKEPSPTTPK- 480

Qy     304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
           |||  |
Db     480 ----EPAPTTT---------------------------------------------------- 487

Qy     364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                    :|      | |  |    || | |  |     |
Db     487 ---------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 519

Qy     423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
           |  :    | |                                :| |    :| |
Db     520 APTT-----PKEP-----------------------APTTPKKPAPTT----------- 540

Qy     483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                ||||     |||     ||||   |:   |  | | |   : |||     ||
Db     540 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 596

Qy     536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
             ||||     ||||    :|||   | ||    ||| |  |: |   |    |||
Db     597 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 654

Qy     590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT-----------VTVPP 634
            |||      ||||     |||:  |    |   ||   || ||             | |
Db     655 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 714

Qy     635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH-------------APNTTAAPVTTTS 679
             ||||   ||| |     ||     || |:                || ||  | :|||
Db     715 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 774

Qy     680 ---APATTPEDDDIDP--PLPNDPINP 701
              || |  |     |  | |  |  |
Db     775 DKPAPTTPKETAPTTPKEPAPTTPKKP 801


Search completed: July 18, 1999, 03:46:46
Job time: 1530 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 07:49:59 ; Search time 21.52 Seconds
(without alignments)
361.351 Million cell updates/sec

Title: US-09-294-663-3
Perfect score: 4328
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 788

Scoring table: BLOSUM62

Searched: 106577 seqs, 9868381 residues

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/PCTUS9_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	313	7.2	907	3	PCT-US95-04611A-19	Sequence 19, Appl
2	308.5	7.1	451	1	US-08-287-001A-2	Sequence 2, Appli
3	308.5	7.1	451	3	PCT-US95-09941-2	Sequence 2, Appli
4	279	6.4	2476	2	US-08-276-967-2	Sequence 2, Appli
5	265.5	6.1	214	1	US-08-217-327-4	Sequence 4, Appli
6	258	6.0	2035	2	US-08-479-537A-2	Sequence 2, Appli
7	258	6.0	1867	2	US-08-479-537A-5	Sequence 5, Appli
8	237	5.5	486	1	US-08-450-360-2	Sequence 2, Appli
9	235.5	5.4	1537	1	US-08-325-267A-2	Sequence 2, Appli
10	219.5	5.1	629	3	PCT-US95-07721-6	Sequence 6, Appli
11	218	5.0	722	1	US-08-347-718B-1	Sequence 1, Appli
12	218	5.0	742	1	US-08-347-718B-2	Sequence 2, Appli
13	218	5.0	745	2	US-08-445-050-2	Sequence 2, Appli
14	218	5.0	722	2	US-08-445-050-3	Sequence 3, Appli
15	218	5.0	722	2	US-08-445-050-7	Sequence 7, Appli
16	218	5.0	722	2	US-08-482-262-1	Sequence 1, Appli
17	218	5.0	742	2	US-08-482-262-2	Sequence 2, Appli
18	218	5.0	745	2	US-08-204-691-2	Sequence 2, Appli
19	218	5.0	722	2	US-08-204-691-3	Sequence 3, Appli
20	218	5.0	722	2	US-08-204-691-7	Sequence 7, Appli
21	217.5	5.0	357	1	US-08-078-683A-8	Sequence 8, Appli
22	216	5.0	2035	1	US-08-046-585-5	Sequence 5, Appli
23	216	5.0	2035	1	US-08-393-703-5	Sequence 5, Appli
24	216	5.0	2035	3	PCT-US93-11721-5	Sequence 5, Appli
25	213	4.9	447	1	US-08-450-360-4	Sequence 4, Appli
26	211	4.9	895	1	US-08-123-161A-8	Sequence 8, Appli
27	211	4.9	895	1	US-08-483-278-8	Sequence 8, Appli
28	206.5	4.8	806	1	US-08-270-076A-11	Sequence 11, Appl
29	206.5	4.8	985	3	PCT-US96-03916-6	Sequence 6, Appli
30	206.5	4.8	985	3	PCT-US96-03916-66	Sequence 66, Appl
31	205	4.7	862	1	US-08-325-267A-4	Sequence 4, Appli
32	199	4.6	1848	3	PCT-US95-10661A-6	Sequence 6, Appli
33	197.5	4.6	921	1	US-08-188-582-2	Sequence 2, Appli
34	197.5	4.6	921	1	US-08-646-715-2	Sequence 2, Appli
35	195.5	4.5	1706	2	US-08-459-568-2	Sequence 2, Appli
36	195.5	4.5	1706	2	US-08-399-411-2	Sequence 2, Appli
37	192.5	4.4	2441	2	US-08-194-468-2	Sequence 2, Appli
38	192.5	4.4	1719	2	US-08-459-568-4	Sequence 4, Appli
39	192.5	4.4	412	2	US-08-477-254A-4	Sequence 4, Appli
40	192.5	4.4	1719	2	US-08-399-411-4	Sequence 4, Appli
41	192.5	4.4	412	2	US-08-472-576B-4	Sequence 4, Appli
42	192.5	4.4	412	2	US-08-428-734B-4	Sequence 4, Appli
43	190.5	4.4	1481	2	US-08-616-844-40	Sequence 40, Appl
44	190.5	4.4	1481	2	US-08-599-654-40	Sequence 40, Appl
45	187.5	4.3	2414	1	US-08-227-536-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT   1
PCT-US95-04611A-19
; Sequence 19, Application PC/TUS9504611A
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard and Jackman, Winthrop, T.
;    TITLE OF INVENTION:  Non Splicing Variants of gp350/220
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward Castro Huddleson & Tatum
;      STREET:  5 Palo Alto Square
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04611A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/229,291
;      FILING DATE:  April 18, 1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  AVIR-003/00US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-843-5163
;      TELEFAX:  415-857-0663
;      TELEX:  380816 CooleyPA
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  907 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-04611A-19


  Query Match             7.2%;  Score 313;  DB 3;  Length 907;
  Best Local Similarity   22.5%;  Pred. No. 4.7e-14;
  Matches  168;  Conservative   65;  Mismatches  221;  Indels  294;  Gaps   32;

Qy      43 WLLPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGC-TLPGPPAET-- 100
           |  |:  :    : |:: |    |  |       |::       ::|  | |     |
Db     352 WAWPNNTETD--FKCKWTLTSGTPSGC-ENISGAFASNRTFDITVSGLGTAPKTLIITRT 408

Qy     100 -TQAPATTQAPTTTQAPTTTTQAPTTTTQA---PTTTTQAPTTTQAPTTTQAPTTTQAPT 155
            | |  ||     ::|| :|| :||  |     | |||  |::|  ||   || :|  ||
Db     409 ATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPAST-GPT 467

Qy     156 TTTQAPTTTTQAPTTTTQAPTT----------TQAP--TTTQAPTTTQAP--TTITQAA 200
           :|   |: | | ||:  :| |           ::||  |:: :| ||  |  |: | |
Db     468 VSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTSSTSPVTTPTPNATSPTPAV 527

Qy     201 TTPAATTPAAT--TPAATTPAATTPAATTP--GVPAPTSAPVWPPICELLPNGCPADFDI 256
           |||   || ||  ||| |||   || ||:|  |  :|||   |      ||
```

```
Db    528 TTP---TPNATSPTPAVTTP---TPNATSPTLGKTSPTSAVTTP-----TPN-------- 569

Qy    257 HLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEG 316
                                                 ||      |: ||   ||
Db    569 ------------------------------------ATSPTL----GKTSPTSAVT-- 585

Qy    317 NEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQ 376
                                      |  |
Db    585 -------------------------TPTPN------------------------------ 590

Qy    377 CDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESD 436
                 | |               |   ||  :   | ||
Db    590 --------ATSP-------------TLGKTSPTSAVT-TPTPNAT--------------- 613

Qy    437 CGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPT 496
                 | |:    |                          : :: |      :||   |
Db    613 ---------GPTVGETSP----------------------QANATNHTLGGTSPTPVVT 640

Qy    497 AAPTAAPTAA------------------PTAAPSTVVPPATPPATAAPVPPTTAIPT--- 536
          : |  | :|                   |:: | |: |  :  :|:     |:| ||
Db    641 SQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSDNSTSHMPLLTSAHPTGGE 700

Qy    536 ----PAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVP-EIPITVTSAP-----T 585
              | :  |   :|::|| |   |    : |  : |:  | |: :|  : |
Db    701 NITQVTPASISTHHVSTSSPE-PRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNATSP 759

Qy    586 AAPTAAPTAAPTAAPTTAVPEIPT-----------TVTSPPT---AAPTTAAPAPNTTVT 631
           ||:   || ||   |       |            | | | |      ||  |  | |
Db    760 QAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRPRYNATTY 819

Qy    632 VPP------------TAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVT--- 677
          :||            |: | | | |   || ||||: |  :  ::      : | :|
Db    820 LPPSTSSKLRPRWTFTSPPVTTAQA---TVPVPPTSQPRFSNLSMLVLQWASLAVLTLLL 876

Qy    677 -------------TTSAPATTPEDDDID 691
                       :||   |||  || :
Db    877 LLVMADCAFRRNLSTSHTYTTPPYDDAE 904
```

```
RESULT   2
US-08-287-001A-2
; Sequence 2, Application US/08287001A
; Patent No. 5622861
;  GENERAL INFORMATION:
;    APPLICANT:  KAPLAN, GERARDO
;    APPLICANT:  FEINSTONE, STEPHEN M.
;    TITLE OF INVENTION:  HEPATITIS A VIRUS RECEPTOR AND METHODS
;    TITLE OF INVENTION:  OF USE
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg, 127 Peachtree
;      STREET:  Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/287,001A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spratt, Gwedolyn D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  1414.621
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  451 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-287-001A-2


  Query Match             7.1%;  Score 308.5;  DB 1;  Length 451;
  Best Local Similarity   41.6%;  Pred. No. 4.3e-14;
  Matches   87;  Conservative    4;  Mismatches   53;  Indels   65;  Gaps   10;

Qy     94 GPPAET----------TQAPATTQAPTTTQAPTTTTQAPTTTTQAPTT----TTQAPTTT 139
          |||  |          |  | ||  ||||  |||||  |||||   ||    ||  ||||
Db    125 GPPRVTTPIVRTVRTSTTVPTTTTLPTTTTLPTTTT-LPTTTTLPMTTTLPMTTTLPTTT 183

Qy    140 QAPTTTQAPTTTQAPTT---TTQAPTTTTQAPTTTT-----QAPTTTQAPTTTQAPTTTQ 191
            ||||  ||||  |||    ||  ||| |  |||||       | ||  ||||  ||||
Db    184 TVPTTTTLPTTTTLPTTLPMTTTLPTTRT-LPTTTTLPTTMTLPMTTTLPTTTTLPTTTT 242

Qy    192 APTTITQAATTPAATT-------PAATT-PAATTPAATT----------------PAAT 226
            |||      |  |  ||       |  || |  ||   ||                | ||
Db    243 LPTTTLPTMTLPTTTTLPTMMTLPTTTTLPTTTTLPTTTMVSTFVPPTPLPMQNHEPVAT 302

Qy    227 TPGVP----------------APTSAPVW 239
          :|  |                |||:|::
Db    303 SPSSPQPAETHPVTLLGATRTQPTSSPLY 331


RESULT   3
PCT-US95-09941-2
; Sequence 2, Application PC/TUS9509941
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  HEPATITIS A VIRUS RECEPTOR AND METHODS
;    TITLE OF INVENTION:  OF USE
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg, 127 Peachtree
;      STREET:  Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/09941
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/287,001
;      FILING DATE:  5 AUG 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spratt, Gwedolyn D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  1414.621
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  451 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-09941-2
```

```
  Query Match             7.1%;  Score 308.5;  DB 3;  Length 451;
  Best Local Similarity   41.6%;  Pred. No. 4.3e-14;
  Matches   87;  Conservative    4;  Mismatches   53;  Indels   65;  Gaps   10;

Qy      94 GPPAET----------TQAPATTQAPTTTQAPTTTTQAPTTTTQAPTT----TTQAPTTT 139
           |||  |          |  | ||  ||||  |||||  |||||   ||    ||  ||||
Db     125 GPPRVTTPIVRTVRTSTTVPTTTTLPTTTTLPTTTT-LPTTTTLPMTTTLPMTTTLPTTT 183

Qy     140 QAPTTTQAPTTTQAPTT---TTQAPTTTTQAPTTTT-----QAPTTTQAPTTTQAPTTTQ 191
           ||||  ||||  |||   ||  ||| |  |||||       | ||  ||||  ||||
Db     184 TVPTTTTLPTTTTLPTTLPMTTTLPTTRT-LPTTTTLPTTMTLPMTTTLPTTTTLPTTTT 242

Qy     192 APTTITQAATTPAATT-------PAATT-PAATTPAATT-----------------PAAT 226
           |||     | |  ||       |  || |  ||   ||                 | ||
Db     243 LPTTTLPTMTLPTTTTLPTMMTLPTTTTLPTTTTLPTTTMVSTFVPPTPLPMQNHEPVAT 302

Qy     227 TPGVP----------------APTSAPVW 239
           :|  |                |||:|::
Db     303 SPSSPQPAETHPVTLLGATRTQPTSSPLY 331


RESULT   4
US-08-276-967-2
; Sequence 2, Application US/08276967
; Patent No. 5851817
;  GENERAL INFORMATION:
;    APPLICANT:  Hardy, Daniel M.
;    APPLICANT:  Garbers, David L.
;    TITLE OF INVENTION:  Species-Specific Egg-Binding Proteins of
;    TITLE OF INVENTION:  Sperm
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P. O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210-4433
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,967
;      FILING DATE:  Submitted Herewith
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitchell, Barbara S.
;      REGISTRATION NUMBER:  33,928
;      REFERENCE/DOCKET NUMBER:  UTSD:418\KIT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-787-1400
;      TELEFAX:  713-789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2476 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-276-967-2


  Query Match             6.4%;  Score 279;  DB 2;  Length 2476;
  Best Local Similarity   21.2%;  Pred. No. 3.1e-11;
  Matches  166;  Conservative   43;  Mismatches  234;  Indels  340;  Gaps   37;

Qy      34 PDCPPAEQHWLLPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLP 93
           | | ||          |  | |  |||          ||: :       | |
Db     217 PFCAPA---------VICVTFTYHMYGL--------GQGTKLRL---------LLGSPAG 250

Qy      94 GPPAET-----TQAP----ATTQAPTTTQAPTT------------------------TT 119
           ||:        |:|     :   |:  | |                          |
Db     251 SPPSSLWERVGPQSPEWLNTSVTIPSGHQQPMQLIFEAVRGTNTAFVVALGFVLINHGTC 310

Qy     120 QAPTTT---TQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPT 176
           : |: |   |: |   |: ||      |       |: |    : |   |: ||  |: ||
Db     311 RGPSETSVSTEKPVAPTEKPTVPSEIYT----IPTEKPMVHMEKPIVHTEKPTVPTEKPT 366

Qy     177 --TTQAPTTTQAPTTTQAPTTITQAATTPA--ATTP---AATTPAATTPAATTPAATTPG 229
           | ::   |: ||  : ||   :  | ||   |||    | |   |     | : |
Db     367 IPTEKSTVPTKKPTVFKEPTLPPEGPTVPAERPTTPPEGPAVPPKGPTVLTEWPTSHTEK 426

Qy     230 VPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFN 289
           |  |: |     :|   |       ::|
Db     427 STVHTEKPILPTGKSTIPTEKP-------MVP---------------------------- 452

Qy     290 PYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCD 349
           :|  :|
Db     452 --TKRTTTP------------------------------------------------- 459

Qy     350 KYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPT 409
                   ||                        |:|  :                ||
Db     459 ----------TER----------------------TTIPAEK---------------PT 470

Qy     410 EPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESP 469
           |||   :|                                                |:    |
Db     471 VPIEKPMVP-------------------------------------------TERTTIP 486

Qy     470 VTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPP 529
                     |  ||  ||  ||   |  ||  |  :||      |  |: |
Db     487 T--------------ERTTIPTEKPT-VPTEKLT-VPTEKP--IVP------TEKPIVP 521

Qy     530 TT--AIPTPAPTAAP--TAAP---TTAAPESPTTVTVPPTAAPTAAPT--TAVPEIP--- 578
           |    |||   |      |  |   ||   | ||  |  | : ||  ||  |  | ||
Db     522 TEKHTIPTEKLTVLTERTTTPTERTTIPTEKPTVPTEKP-SVPTEKPTVPTEEPTIPTEK 580

Qy     578 ITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPT--TAAPAPNTTVTVPPT 635
           :|| :  |  ||   |  ||   ||  | | ||  :  |  ||  |  |   ||:    |
Db     581 LTVPTERTTTPTKR-TTTPTIRTTT--PTIRTTTPTERTTTPTIRTTTPTERTTIPTKKT 637

Qy     636 AAPT--TAAPAPNTTVTVPPTAAPTAAP--PTVAHAPNTTAAPVTTTSAPATTPEDDDID 691
           ||  |  |   |     |  :||  |  |     |:|:|    ||  | ||
Db     638 TVPTEKTIIPTERTIAPTTPQPSPTLVPTQPAAVVMPSTSA----TTVTPRTT------- 687

Qy     692 PPLPNDPINPCVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNV 751
             : : | |   | |                          |     |     |  |:  |
Db     687 --IASCPPNAHFERC----------------------ACPVSCQ--SPTPNCELFCKP 718

Qy     752 GCV 754
           |||
Db     719 GCV 721


RESULT   5
US-08-217-327-4
; Sequence 4, Application US/08217327
; Patent No. 5474925
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E
;    APPLICANT:  Barton, Kenneth A
;    TITLE OF INVENTION:  Immobilized Proteins in Cotton Fiber
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O. Box 2113
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  USA
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/217,327
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/812,233
;      FILING DATE:  19-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990831
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  214 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-217-327-4


  Query Match             6.1%;  Score 265.5;  DB 1;  Length 214;
  Best Local Similarity   37.1%;  Pred. No. 1.6e-11;
  Matches   72;  Conservative   15;  Mismatches   62;  Indels   45;  Gaps   10;

Qy     502 APTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPP 561
           ||:  ||: |:|  |||: |      ||||  | | ||| |           |   | ||
Db      25 APSNPPTSTPATPTPPASTP------PPTTQAP-PTPTATP-----------PPVSTPPP 66

Qy     562 TAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTT 621
           |::|   | || |  |    |  |::| | | |   | |:|  | |       ||| | |
Db      67 TSSP--PPVTASP--PPVSTPPPSSPPPATP---PPASPPPATP----PPASPPPATPPP 115

Qy     622 AAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAP 681
           |:| | |    || | |  | | | |     |  || |:||    | :    |:|:  ||
Db     116 ASPPPAT----PPPATPPPATPPPAT-----PPPAPLASPPATVPAISPVQTPLTSPPAP 166

Qy     682 ATTPEDDDIDPPLP 695
           |       : | |
Db     167 PT-------EAPAP 173


RESULT   6
US-08-479-537A-2
; Sequence 2, Application US/08479537A
; Patent No. 5861381
;  GENERAL INFORMATION:
;    APPLICANT:  CHAMBON, Pierre
;    APPLICANT:  KIENY, Marie-Paule
;    APPLICANT:  LATHE, Richard
;    APPLICANT:  HAREUVENI, Mara
;    TITLE OF INVENTION:  PHARMACEUTICAL COMPOSITION FOR THE
;    TITLE OF INVENTION:  TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,537A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 90/13101
;      FILING DATE:  23-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR91/00835
;      FILING DATE:  23-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/039,320
;      FILING DATE:  04-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/403,576
;      FILING DATE:  14-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  017753-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  128..1899
;      OTHER INFORMATION:  /note= "The amino acids spanning
;      OTHER INFORMATION:  128 to 1899 constitute a repeated region wherein the repea
;      OTHER INFORMATION:  20 amino acids, 17 of which are fixed. The number of such
;      OTHER INFORMATION:  repeats varies from 1 to 40."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  134
;      OTHER INFORMATION:  /note= "Amino acid 134 is X1 = Xaa
;      OTHER INFORMATION:  Xaa Xaa which is the codon for Pro or Ala wherein Pro = CC
;      OTHER INFORMATION:  CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  144
;      OTHER INFORMATION:  /note= "Amino acid 144 is Y = Xaa
;      OTHER INFORMATION:  which is the codon for Thr or Asn wherein Thr = ACT, ACC,
;      OTHER INFORMATION:  or ACG; and Asn = AAT or AAC."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  147
;      OTHER INFORMATION:  /note= "Amino acid 147 is X2 = Xaa
;      OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC,
;      OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  1..21
;      OTHER INFORMATION:  /note= "Amino acids 1 to 21 are a
;      OTHER INFORMATION:  21 amino acid precursor sequence."
US-08-479-537A-2


  Query Match             6.0%;  Score 258;  DB 2;  Length 2035;
  Best Local Similarity   23.4%;  Pred. No. 6.6e-10;
  Matches  160;  Conservative   43;  Mismatches  280;  Indels  202;  Gaps   34;

Qy      93 PGPPAETTQAPATTQAPTTTQAPTT-------TTQAPTTTTQAPTTTTQAPTTTQAPTTT 145
           ||  : |||    | || |  |  :        |  | |     :||  |   | ||
Db      66 PGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 125

Qy     146 QAPTTTQAPTT--TTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAP-----TTITQ 198
            || :| ||     | ||        :|   |   | ||     | :| ||     |:
Db     126 PAPGST-APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGST-APXAHGVTSAPD 183

Qy     199 AATTPAATTPAA----------TTPAATTPAA----------TTPAATTPGVPAPTSAPV 238
```

```
Db    184 XRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPD 243

Qy    239 WPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSP 298

Db    244 XRPX-----PGSTAPX-----------------AHGVT-----------------SAP 262

Qy    299 ANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGN 358

Db    263 DXRPXPGSTAPX-------------------------------------------AHG- 278

Qy    359 LVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP-----TEPIE 413

Db    278 -------------VTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPD 323

Qy    414 WEPLPNG-CPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCPGNLHFSPATQSCESP 469

Db    324 XRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS----------TAP 373

Qy    470 VTAGCQVFECDSDNQCTSTAAPTA-APTAAPTAAP---TAAPTAAPSTVVPPATP-PATA 524

Db    374 XAHG--VTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGST 431

Qy    525 AP-VPPTTAIP--TPAP-TAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEI-PIT 579

Db    432 APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPX---PGSTAPXAHGVTSAPDXRPXP 488

Qy    580 VTSAP-----TAAPTAAP---TAAPTAAPTTAVPEI-PTTVTSPPTAAPTTAA----PAP 626

Db    489 GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXP 548

Qy    627 NTT------VTVPPTAAPTTAAPAPNT-TVTVPPTAAP---TAAPPT--VAHAPNTTAAP 674

Db    549 GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXP 608

Qy    675 VTT-------TSAPATTPEDDDIDP 692

Db    609 GSTAPXAHGVTSAPDXRPXPGSTAP 633


RESULT   7
US-08-479-537A-5
; Sequence 5, Application US/08479537A
; Patent No. 5861381
;  GENERAL INFORMATION:
;    APPLICANT:  CHAMBON, Pierre
;    APPLICANT:  KIENY, Marie-Paule
;    APPLICANT:  LATHE, Richard
;    APPLICANT:  HAREUVENI, Mara
;    TITLE OF INVENTION:  PHARMACEUTICAL COMPOSITION FOR THE
;    TITLE OF INVENTION:  TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,537A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 90/13101
;      FILING DATE:  23-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR91/00835
;      FILING DATE:  23-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/039,320
;      FILING DATE:  04-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/403,576
;      FILING DATE:  14-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  017753-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1867 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  128..1727
;      OTHER INFORMATION:  /note= "The amino acids spanning
;      OTHER INFORMATION:  128 to 1727 constitute a repeated region wherein the repea
;      OTHER INFORMATION:  20 amino acids, 17 of which are fixed.  The number of such
;      OTHER INFORMATION:  repeats varies from 1 to 40."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  134
;      OTHER INFORMATION:  /note= "Amino acid 134 is X1 = Xaa
;      OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC,
;      OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  144
;      OTHER INFORMATION:  /note= "Amino acid 144 is Y = Xaa
;      OTHER INFORMATION:  which is the codon for Thr or Asn wherein Thr = ACT, ACC,
;      OTHER INFORMATION:  or ACG; and Asn = AAT or AAC."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  147
;      OTHER INFORMATION:  /note= "Amino acid 147 is X2 = Xaa
;      OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC,
;      OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  1..21
;      OTHER INFORMATION:  /note= "Amino acids 1 to 21 are a
;      OTHER INFORMATION:  21 amino acid precursor sequence."
US-08-479-537A-5


  Query Match             6.0%;  Score 258;  DB 2;  Length 1867;
  Best Local Similarity   23.4%;  Pred. No. 6e-10;
  Matches  160;  Conservative   43;  Mismatches  280;  Indels  202;  Gaps   34;

Qy     93 PGPPAETTQAPATTQAPTTTQAPTT-------TTQAPTTTTQAPTTTTQAPTTTQAPTTT 145

Db     66 PGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 125

Qy    146 QAPTTTQAPTT--TTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAP-----TTITQ 198

Db    126 PAPGST-APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGST-APXAHGVTSAPD 183

Qy    199 AATTPAATTPAA----------TTPAATTPAA----------TTPAATTPGVPAPTSAPV 238

Db    184 XRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPD 243

Qy    239 WPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSP 298

Db    244 XRPX-----PGSTAPX-----------------AHGVT-----------------SAP 262
```

```
Qy    299 ANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGN 358
Db    263 DXRPXPGSTAPX-------------------------------------------AHG- 278

Qy    359 LVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP-----TEPIE 413
Db    278 --------------VTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPD 323

Qy    414 WEPLPNG-CPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCPGNLHFSPATQSCESP 469
Db    324 XRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS----------TAP 373

Qy    470 VTAGCQVFECDSDNQCTSTAAPTA-APTAAPTAAP---TAAPTAAPSTVVPPATP-PATA 524
Db    374 XAHG--VTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGST 431

Qy    525 AP-VPPTTAIP--TPAP-TAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEI-PIT 579
Db    432 APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPX---PGSTAPXAHGVTSAPDXRPXP 488

Qy    580 VTSAP-----TAAPTAAP---TAAPTAAPTTAVPEI-PTTVTSPPTAAPTTAA----PAP 626
Db    489 GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXP 548

Qy    627 NTT------VTVPPTAAPTTAAPAPNT-TVTVPPTAAP---TAAPPT--VAHAPNTTAAP 674
Db    549 GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXP 608

Qy    675 VTT-------TSAPATTPEDDDIDP 692
Db    609 GSTAPXAHGVTSAPDXRPXPGSTAP 633
```

```
RESULT   8
US-08-450-360-2
; Sequence 2, Application US/08450360
; Patent No. 5656457
;  GENERAL INFORMATION:
;    APPLICANT:  Parkes, Deborah Lynn
;    APPLICANT:  Coates, Stephen Ralph
;    TITLE OF INVENTION:  Herpes Simplex Virus Type 2-Glycoprotein G
;    TITLE OF INVENTION:  Proteins and Polypeptides
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Leona L. Lauder
;      STREET:  6 Mariposa Court
;      CITY:  Tiburon
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94920
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/450,360
;      FILING DATE:  25-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/129,021
;      FILING DATE:  29-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lauder, Leona L.
;      REGISTRATION NUMBER:  30,863
;      REFERENCE/DOCKET NUMBER:  D-0012.2A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-435-2034
;      TELEFAX:  415-435-0727
;  GENERAL INFORMATION:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-450-360-2


  Query Match             5.5%;  Score 237;  DB 1;  Length 486;
  Best Local Similarity   30.9%;  Pred. No. 3.5e-09;
  Matches   92;  Conservative   23;  Mismatches   95;  Indels   88;  Gaps   14;

Qy    408 PTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQS-- 466
Db    256 PTEDVGVLP----------PHWAPGALDDGPY----------APFPPRPRFRRALRTDP 294

Qy    466 ------CESPVTAGCQVFECDSDNQCTSTAAPTAAP----------TAAPTAAPTAAPTA 509
Db    295 EGVDPDVRAPRT-GRRLMALTED---TSSDSPTSAPEKTPLPVSATAMAPSVDPSAEPT- 350

Qy    510 APSTVVPP---ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPT 566
Db    350 APATTTPPDEMATQAATVAVTPEETAVASP---------PATASVES----------SPL 390

Qy    567 AAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAP 626
Db    391 PAAAAATPGAGHTNTSSASAAKTPPTTPAPTTPPPTSTHATP----RPTTPGPQTTPPGP 446

Qy    627 NTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATT 684
Db    447 ATPGPVGASAAPTADSPL------------TASPP--ATAPGPSAANV---SVAATT 486
```

```
RESULT   9
US-08-325-267A-2
; Sequence 2, Application US/08325267A
; Patent No. 5585271
;  GENERAL INFORMATION:
;    APPLICANT:  WATARI, JUNJI
;    APPLICANT:  TAKATA, YOSHIHIRO
;    APPLICANT:  OGAWA, MASAHIRO
;    APPLICANT:  PENTTILA, MERJA
;    APPLICANT:  ONNELA, MAIJA-LEENA
;    APPLICANT:  KERANEN, SIRKKA
;    TITLE OF INVENTION:  YEAST AGGLUTINATION GENES AND YEAST
;    TITLE OF INVENTION:  CONTAINING THEM
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVIS HWY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,267A
;      FILING DATE:  18-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP PCT/JP94/00290
;      FILING DATE:  24-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 38871/1993
;      FILING DATE:  26-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2589-023-0XPCT
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 1537 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-325-267A-2


  Query Match           5.4%;  Score 235.5;  DB 1;  Length 1537;
  Best Local Similarity  23.2%;  Pred. No. 1.6e-08;
  Matches  166;  Conservative   45;  Mismatches  291;  Indels  213;  Gaps   31;

Qy      88 AGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQA 147
           : ||:| |   :  | :||   ||| | | |   |:|     | |    | :    :
Db     261 SNCTVPDP---SNYAVSTT---TTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRT 314

Qy     148 PTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAATTPAATT 207
           |||     |||:   :|  : :|    | |    | :    : ||| | | ||   |
Db     315 PTTASTIITTTTEPWNSTFTSTSTELTTVTGTNGVRTDETIIVIRTPTTATTAITT---TE 371

Qy     208 PAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCN 267
           |  :|  :|:   ||   |                     || | |  | ::
Db     372 PWNSTFTSTSTELTTVTGT--------------------NGLPTDETIIVI-------- 403

Qy     268 LFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDL 327
                     | |              | |       :    ||
Db     403 -------------RTPTTA---TTAMTTTQPWNDTFTSTSTEMTTVTG------------ 435

Qy     328 LDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTL 387
             || | :  |                 :| |     |     |  |        |
Db     435 -TNGLPTDETI-------------------IVIRTPTTATTAMTTTQPWDD------TF 467

Qy     388 PGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQ 447
              :|   |     ||              || | | :|
Db     468 TSTSTEMTTV-----TG-------------TNGLPTDETI-------------------- 490

Qy     448 TIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPT-------AAPT 500
           : |        :| | :     |        :  : |:       ||          ||
Db     490 IVIR--------TPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPT 541

Qy     501 AAPTAAPTAAP--STVVPPATPPATAAPVPPTTAIPTP-------APTAAPTAAPTTAAP 551
           | ||  |  |    |    :|   |   |  |  :||         || | ||  ||
Db     542 TATTAMTTTQPWNDTFTSTSTEMTT---VTGTNGLPTDETIIVIRTPTTATTAITTTEPW 598

Qy     552 ESPTTVTVPPTAAPTAAPTTAVP-EIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTT 610
           |  | |   |  .|   |  :| :  | |   || | ||  |  |      |:     ||
Db     599 NS--TFTSTSTELTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTT 656

Qy     611 VTSP---PT-------AAPTTAAPAPNT----------TVTVPPTAAPTTAAPAPNTTVT 650
           ||   ||        .||||  |  |           | |   |   |   |   | :
Db     657 VTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIV 716

Qy     651 V-PPTAAPTAAPPTVAHAP-----NTTAAPVTTTSAPATTPEDDD---IDPPLPNDPINP 701
           :  || | ||   |    |      :|:  :|| :     | |:    |  |     |:
Db     717 IRTPTTATTAMTTT---QPWNDTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLIST 773

Qy     702 CVEECNVLPW----AHADCDKYWVCDGNNQ-----VLVV---CSEGLQFNPTTKT 744
             |    ||          :   |   | |     |:|:    ||||    || |
Db     774 TTE-----PWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGL---VTTTT 820


RESULT  10
PCT-US95-07721-6
; Sequence 6, Application PC/TUS9507721
;  GENERAL INFORMATION:
;    APPLICANT:  Massachusetts Institute of Technology
;    TITLE OF INVENTION:  Class BI and CI Scavenger Receptors
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/07721
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  MIT6620
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  629 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    FEATURE:
;      NAME/KEY:  misc. feature
;      LOCATION:  1..629
;      OTHER INFORMATION:  /Function = "Amino acid sequence for the
;      OTHER INFORMATION:  Drosophila Melanogaster Scavenger Receptor
;      OTHER INFORMATION:  Class CI."
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  30..353
;      OTHER INFORMATION:  /note= "Positions 30-32, 90-92,
;      OTHER INFORMATION:  129-131, 180-182, 253-255 and 351-353
;      OTHER INFORMATION:  represent potential N-glycosylation
;      OTHER INFORMATION:  sites."
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  1..20
;      OTHER INFORMATION:  /note= "Amino acids 1-20 represent
;      OTHER INFORMATION:  a putative signal sequence."
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  21..74
;      OTHER INFORMATION:  /note= "Amino acids 21-74 represent
;      OTHER INFORMATION:  complement control protein domain
;      OTHER INFORMATION:  number 1."
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  75..127
;      OTHER INFORMATION:  /note= "Amino acids 75-127
;      OTHER INFORMATION:  represent complement control protein
;      OTHER INFORMATION:  domain number 2."
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  128..312
;      OTHER INFORMATION:  /note= "Amino acids 128-312
;      OTHER INFORMATION:  represent an MAM domain."
;    FEATURE:
;      NAME/KEY:  Disulfide-bond
;      LOCATION:  22..381
;      OTHER INFORMATION:  /note= "The cysteines at positions
;      OTHER INFORMATION:  22, 45, 59, 72, 77, 99, 113, 125, 136, 144, 216,
;      OTHER INFORMATION:  217, 254, 310, 339, 343, 361, 363, 367, 373, 374
```

```
;      OTHER INFORMATION:  and 381 represent potential disulfide linkages."
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  338..381
;      OTHER INFORMATION:  /note= "Amino acids 338-381
;      OTHER INFORMATION:  represent a somatomedin B domain."
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  387..514
;      OTHER INFORMATION:  /note= "Amino acids 387-514
;      OTHER INFORMATION:  represent a mucin-like potential
;      OTHER INFORMATION:  O-linked glycosylation region."
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  544..564
;      OTHER INFORMATION:  /note= "Amino acids 544-565
;      OTHER INFORMATION:  represent a putative TM domain."
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  565..629
;      OTHER INFORMATION:  /note= "Amino acids 565-629
;      OTHER INFORMATION:  represent a putative cytoplasmic
;      OTHER INFORMATION:  domain."
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  576..602
;      OTHER INFORMATION:  /note= "Amino acids 576-579 and
;      OTHER INFORMATION:  599-602 represent casein kinase II
;      OTHER INFORMATION:  sites."
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  578..592
;      OTHER INFORMATION:  /note= "Amino acids 578-580 and
;      OTHER INFORMATION:  590-592 represent protein kinase C
;      OTHER INFORMATION:  sites."
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  596..599
;      OTHER INFORMATION:  /note= "Amino acids 596-599
;      OTHER INFORMATION:  represent a cAMP protein kinase site."
PCT-US95-07721-6


  Query Match             5.1%;  Score 219.5;  DB 3;  Length 629;
  Best Local Similarity   41.4%;  Pred. No. 7.3e-08;
  Matches   60;  Conservative   16;  Mismatches   42;  Indels   27;  Gaps    6;

Qy      78 SAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPT 137
           | :||    |  |          |:   ||     :  | | |   |:||:  ||||   |
Db     369 SDETCCLNYLEEC---------TKELLTTTEDDISSLPPTVTSTSTSTTRKSTTTTTST 419

Qy     138 TTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTIT 197
           || : |||: ||||          ||||:|  |||:  |||: ||||   :||  ||| |
Db     420 TTTSTTTTKRPTTT----------TTTTKA--TTTKRTTTTKKPTTT---STTPKPTTTT 464

Qy     198 QAATTPAATTPAATTPAATTPAATT 222
             :||| :|| : |:  :|||  ||
Db     465 --STTPKSTT-STTSTTSTTPTTTT 486


RESULT  11
US-08-347-718B-1
; Sequence 1, Application US/08347718B
; Patent No. 5696087
;  GENERAL INFORMATION:
;    APPLICANT:  Wang, Chu-San
;    APPLICANT:  Tang, Jordan J.N
;    TITLE OF INVENTION:  METHOD FOR REDUCING INTESTINAL ABSORPTION OF
;    TITLE OF INVENTION:  CHOLESTEROL
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/347,718B
;      FILING DATE:  December 1, 1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  OMRF150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  722 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
US-08-347-718B-1


  Query Match             5.0%;  Score 218;  DB 1;  Length 722;
  Best Local Similarity   24.9%;  Pred. No. 1.1e-07;
  Matches  117;  Conservative   44;  Mismatches  188;  Indels  120;  Gaps   23;

Qy     304 DGEISPAPPVTEGNEDEDIDI---GDLLDNGCPANFEIDWLLPHGNRCDK------YYQC 354
           ||:  || |:       |||     : :|    |: :    :|  |: :|      :|:
Db     292 DGDFIPADPINLYANAADIDYIAGTNNMDGHIFASID----MPAINKGNKKVTEEDFYKL 347

Qy     355 VHGNLVERRC-GAGTHFSFELQQCDHIELVGCTLPGGESEE---VDVDEDACTGWYCPTE 410
           |    : :   || | |    :            |  |:::   || : |    :  |||
Db     348 VSEFTITKGLRGAKTTFDVYTES-------WAQDPSQENKKKTVVDFETDVL--FLVPTE 398

Qy     411 PIEWEPLPNGCPADFSIDHLLPHES----------DCGQYLQCVHGQTIARPCPGNLHF 459
           :   |   | :  :| | |           |    :| | |:  | |      :
Db     399 IALAQHRANAKSAK-TYAYLFSHPSRMPVYPKWVGADHADDIQYVFGKPFATP----TGY 453

Qy     460 SPATQSCESPVTAGCQVFECDSDNQCTSTAAPT-------------------------- 493
             |  ::    : |    |    |     :| ||
Db     454 RPQDRTVSKAMIAYWTNFAKTGDPNMGDSAVPTHWEPYTTENSGYLEITKKMGSSSMKRS 513

Qy     493 --------------AAPTAAPTAAPTAAPTA-APSTVVPPATPPATAAPVPPTTAIPTPA 537
                         | ||     |    ||  : :| ||| |  :  ||||||
Db     514 LRTNFLRYWTLTYLALPTVTDQEATPVPPTGDSEATPVPP-TGDSETAPVPPT------G 566

Qy     538 PTAAPTAAPTTAAPESPTTVTVPPT---AAPTAAPT--TAVPEIPITVTSAPTAAPTAAP 592
           : ||   ||  :   |    ||||    ||  ||  :  | :| |  |      |
Db     567 DSGAPPVPPTGDSGAPP----VPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGD 622

Qy     593 TAAPTAAPT---TAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPT---AAPTTAAPAPN 646
           : ||   ||    | |  ||    ||   || : ||     ||||    ||       :
Db     623 SGAPPVPPTGDSGAPPVPPTGDAGPPPVPPTGDSGAP----PVPPTGDSGAPPVTPTGDS 678

Qy     647 TTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLP 695
           |  ||||   : ||| |    :: ||||         | ||  :  :|
Db     679 ETAPVPPT-GDSGAPP-VPPTGDSEAAPV--------PPTDDSKEAQMP 717


RESULT  12
US-08-347-718B-2
; Sequence 2, Application US/08347718B
; Patent No. 5696087
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Wang, Chu-San
;    APPLICANT:  Tang, Jordan J.N
;    TITLE OF INVENTION:  METHOD FOR REDUCING INTESTINAL ABSORPTION OF
;    TITLE OF INVENTION:  CHOLESTEROL
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/347,718B
;      FILING DATE:  December 1, 1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  OMRF150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  742 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  186..187
;      OTHER INFORMATION:  /note= "Position 187 represents a
;      OTHER INFORMATION:  potential N-linked glycosylation site."
;    FEATURE:
;      NAME/KEY:  Modified-site
;      LOCATION:  193..194
;      OTHER INFORMATION:  /note= "The serine at position 194
;      OTHER INFORMATION:  represents an active site serine."
;    FEATURE:
;      NAME/KEY:  misc. feature
;      LOCATION:  1..742
;      OTHER INFORMATION:  /Function = "Amino acid sequence for
;      OTHER INFORMATION:  the Human Milk Bile Salt-activated Lipase."
US-08-347-718B-2


  Query Match             5.0%;  Score 218;  DB 1;  Length 742;
  Best Local Similarity   24.9%;  Pred. No. 1.1e-07;
  Matches  117;  Conservative   44;  Mismatches  188;  Indels  120;  Gaps   23;

Qy     304 DGEISPAPPVTEGNEDEDIDI---GDLLDNGCPANFEIDWLLPHGNRCDK------YYQC 354
           ||: || |:       |||     : :|    |: :    :|  |: :|      :|:
Db     312 DGDFIPADPINLYANAADIDYIAGTNNMDGHIFASID----MPAINKGNKKVTEEDFYKL 367

Qy     355 VHGNLVERRC-GAGTHFSFELQQCDHIELVGCTLPGGESEE---VDVDEDACTGWYCPTE 410
           |    : :   || | |    :            |  |:::   || : |    :  |||
Db     368 VSEFTITKGLRGAKTTFDVYTES-------WAQDPSQENKKKTVVDFETDVL--FLVPTE 418

Qy     411 PIEWEPLPNGCPADFSIDHLLPHES----------DCGQYLQCVHGQTIARPCPGNLHF 459
             |:   |   |  :  :|  | |           |    :| | |:  | |      :
Db     419 IALAQHRANAKSAK-TYAYLFSHPSRMPVYPKWVGADHADDIQYVFGKPFATP----TGY 473

Qy     460 SPATQSCESPVTAGCQVFECDSDNQCTSTAAPT-------------------------- 493
           |  ::    : |    |    |     :| ||
Db     474 RPQDRTVSKAMIAYWTNFAKTGDPNMGDSAVPTHWEPYTTENSGYLEITKKMGSSSMKRS 533

Qy     493 --------------AAPTAAPTAAPTAAPTA-APSTVVPPATPPATAAPVPPTTAIPTPA 537
                         | ||     |    ||  : :| ||| |  :  ||||||
Db     534 LRTNFLRYWTLTYLALPTVTDQEATPVPPTGDSEATPVPP-TGDSETAPVPPT------G 586

Qy     538 PTAAPTAAPTTAAPESPTTVTVPPT---AAPTAAPT--TAVPEIPITVTSAPTAAPTAAP 592
           : ||   ||  :   |    ||||    ||   ||  :  | :| |  |     |
Db     587 DSGAPPVPPTGDSGAPP----VPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGD 642

Qy     593 TAAPTAAPT---TAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPT---AAPTTAAPAPN 646
           : ||   ||    | |  ||    ||   ||  : ||     ||||    ||       :
Db     643 SGAPPVPPTGDSGAPPVPPTGDAGPPPVPPTGDSGAP----PVPPTGDSGAPPVTPTGDS 698

Qy     647 TTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLP 695
             |  ||||   : ||| |    :: ||||          | ||  :  :|
Db     699 ETAPVPPT-GDSGAPP-VPPTGDSEAAPV--------PPTDDSKEAQMP 737


RESULT  13
US-08-445-050-2
; Sequence 2, Application US/08445050
; Patent No. 5763739
;  GENERAL INFORMATION:
;    APPLICANT:  Blaeckberg, Lars
;    APPLICANT:  Edlund, Michael
;    APPLICANT:  Hansson, Lennart
;    APPLICANT:  Hernell, Olle
;    APPLICANT:  Lundberg, Lennart
;    APPLICANT:  Stroemqvist, Mats
;    APPLICANT:  Toernell, Jan
;    TITLE OF INVENTION:  No. 5763739el Polypeptides
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  White & Case
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2787
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/445,050
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/204,691
;      FILING DATE:
;      APPLICATION NUMBER:  SE 9300686-4
;      FILING DATE:  01-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9300722-7
;      FILING DATE:  04-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sterner Ph.D., Richard J
;      REGISTRATION NUMBER:  35,372
;      REFERENCE/DOCKET NUMBER:  1103326-850
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)819-8783
;      TELEFAX:  (212)354-8113
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  745 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-445-050-2
```

```
  Query Match             5.0%;  Score 218;  DB 2;  Length 745;
  Best Local Similarity   24.9%;  Pred. No. 1.1e-07;
  Matches  117;  Conservative   44;  Mismatches  188;  Indels  120;  Gaps   23;

Qy    304 DGEISPAPPVTEGNEDEDIDI---GDLLDNGCPANFEIDWLLPHGNRCDK------YYQC 354
          ||:  || |:       |||     : :|    |: :    :|  |: :|      :|:
Db    315 DGDFIPADPINLYANAADIDYIAGTNNMDGHIFASID----MPAINKGNKKVTEEDFYKL 370

Qy    355 VHGNLVERRC-GAGTHFSFELQQCDHIELVGCTLPGGESEE---VDVDEDACTGWYCPTE 410
          |    : :   || | |    :             |  |:::   || : |    :  |||
Db    371 VSEFTITKGLRGAKTTFDVYTES-------WAQDPSQENKKKTVVDFETDVL--FLVPTE 421

Qy    411 PIEWEPLPNGCPADFSIDHLLPHES-----------DCGQYLQCVHGQTIARPCPGNLHF 459
              :   |   | :  :|  | |           |    :| | |:  | |      :
Db    422 IALAQHRANAKSAK-TYAYLFSHPSRMPVYPKWVGADHADDIQYVFGKPFATP----TGY 476

Qy    460 SPATQSCESPVTAGCQVFECDSDNQCTSTAAPT-------------------------- 493
           |  ::    : |    |    |     :| ||
Db    477 RPQDRTVSKAMIAYWTNFAKTGDPNMGDSAVPTHWEPYTTENSGYLEITKKMGSSSMKRS 536

Qy    493 --------------AAPTAAPTAAPTAAPTA-APSTVVPPATPPATAAPVPPTTAIPTPA 537
                       | ||     |    ||  : :| ||| |  :  ||||||
Db    537 LRTNFLRYWTLTYLALPTVTDQEATPVPPTGDSEATPVPP-TGDSETAPVPPT------G 589

Qy    538 PTAAPTAAPTTAAPESPTTVTVPPT---AAPTAAPT--TAVPEIPITVTSAPTAAPTAAP 592
           : ||   ||  :   |    ||||    ||   ||  :  | :| |  |     |
Db    590 DSGAPPVPPTGDSGAPP----VPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGD 645

Qy    593 TAAPTAAPT---TAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPT---AAPTTAAPAPN 646
          : ||   ||    | |  ||    ||   ||  : ||     ||||    ||      :
Db    646 SGAPPVPPTGDSGAPPVPPTGDAGPPPVPPTGDSGAP----PVPPTGDSGAPPVTPTGDS 701

Qy    647 TTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLP 695
           |  ||||   : ||| |    :: ||||         | ||  :  :|
Db    702 ETAPVPPT-GDSGAPP-VPPTGDSEAAPV--------PPTDDSKEAQMP 740


RESULT  14
US-08-445-050-3
; Sequence 3, Application US/08445050
; Patent No. 5763739
;  GENERAL INFORMATION:
;    APPLICANT:  Blaeckberg, Lars
;    APPLICANT:  Edlund, Michael
;    APPLICANT:  Hansson, Lennart
;    APPLICANT:  Hernell, Olle
;    APPLICANT:  Lundberg, Lennart
;    APPLICANT:  Stroemqvist, Mats
;    APPLICANT:  Toernell, Jan
;    TITLE OF INVENTION:  No. 5763739el Polypeptides
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  White & Case
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2787
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/445,050
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/204,691
;      FILING DATE:
;      APPLICATION NUMBER:  SE 9300686-4
;      FILING DATE:  01-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9300722-7
;      FILING DATE:  04-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sterner Ph.D., Richard J
;      REGISTRATION NUMBER:  35,372
;      REFERENCE/DOCKET NUMBER:  1103326-850
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)819-8783
;      TELEFAX:  (212)354-8113
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  722 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;  MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Mammary gland
US-08-445-050-3


  Query Match             5.0%;  Score 218;  DB 2;  Length 722;
  Best Local Similarity   24.9%;  Pred. No. 1.1e-07;
  Matches  117;  Conservative   44;  Mismatches  188;  Indels  120;  Gaps   23;

Qy    304 DGEISPAPPVTEGNEDEDIDI---GDLLDNGCPANFEIDWLLPHGNRCDK------YYQC 354
          ||:  || |:       |||     : :|    |: :    :|  |: :|      :|:
Db    292 DGDFIPADPINLYANAADIDYIAGTNNMDGHIFASID----MPAINKGNKKVTEEDFYKL 347

Qy    355 VHGNLVERRC-GAGTHFSFELQQCDHIELVGCTLPGGESEE---VDVDEDACTGWYCPTE 410
          |    : :   || | |    :             |  |:::   || : |    :  |||
Db    348 VSEFTITKGLRGAKTTFDVYTES-------WAQDPSQENKKKTVVDFETDVL--FLVPTE 398

Qy    411 PIEWEPLPNGCPADFSIDHLLPHES-----------DCGQYLQCVHGQTIARPCPGNLHF 459
              :   |   | :  :|  | |           |    :| | |:  | |      :
Db    399 IALAQHRANAKSAK-TYAYLFSHPSRMPVYPKWVGADHADDIQYVFGKPFATP----TGY 453

Qy    460 SPATQSCESPVTAGCQVFECDSDNQCTSTAAPT-------------------------- 493
           |  ::    : |    |    |     :| ||
Db    454 RPQDRTVSKAMIAYWTNFAKTGDPNMGDSAVPTHWEPYTTENSGYLEITKKMGSSSMKRS 513

Qy    493 ----------AAPTAAPTAAPTAAPTA-APSTVVPPATPPATAAPVPPTTAIPTPA 537
                     | ||     |    ||  : :| ||| |  :  ||||||
Db    514 LRTNFLRYWTLTYLALPTVTDQEATPVPPTGDSEATPVPP-TGDSETAPVPPT------G 566

Qy    538 PTAAPTAAPTTAAPESPTTVTVPPT---AAPTAAPT--TAVPEIPITVTSAPTAAPTAAP 592
           : ||   ||  :   |    ||||    ||   ||  :  | :| |  |     |
Db    567 DSGAPPVPPTGDSGAPP----VPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGD 622

Qy    593 TAAPTAAPT---TAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPT---AAPTTAAPAPN 646
          : ||   ||    | |  ||    ||   ||  : ||     ||||    ||      :
Db    623 SGAPPVPPTGDSGAPPVPPTGDAGPPPVPPTGDSGAP----PVPPTGDSGAPPVTPTGDS 678

Qy    647 TTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLP 695
           |  ||||   : ||| |    :: ||||         | ||  :  :|
Db    679 ETAPVPPT-GDSGAPP-VPPTGDSEAAPV--------PPTDDSKEAQMP 717


RESULT  15
US-08-445-050-7
; Sequence 7, Application US/08445050
; Patent No. 5763739
;  GENERAL INFORMATION:
;    APPLICANT:  Blaeckberg, Lars
;    APPLICANT:  Edlund, Michael
;    APPLICANT:  Hansson, Lennart
;    APPLICANT:  Hernell, Olle
;    APPLICANT:  Lundberg, Lennart
;    APPLICANT:  Stroemqvist, Mats
;    APPLICANT:  Toernell, Jan
;    TITLE OF INVENTION:  No. 5763739el Polypeptides
```

```
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  White & Case
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2787
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/445,050
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/204,691
;      FILING DATE:
;      APPLICATION NUMBER:  SE 9300686-4
;      FILING DATE:  01-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 9300722-7
;      FILING DATE:  04-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sterner Ph.D., Richard J
;      REGISTRATION NUMBER:  35,372
;      REFERENCE/DOCKET NUMBER:  1103326-850
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)819-8783
;      TELEFAX:  (212)354-8113
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  722 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      TISSUE TYPE:  Mammary gland
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  1..722
;      OTHER INFORMATION:  /label= Variant_N
US-08-445-050-7


  Query Match             5.0%;  Score 218;  DB 2;  Length 722;
  Best Local Similarity   24.9%;  Pred. No. 1.1e-07;
  Matches  117;  Conservative   44;  Mismatches  188;  Indels  120;  Gaps   23;

Qy    304 DGEISPAPPVTEGNEDEDIDI---GDLLDNGCPANFEIDWLLPHGNRCDK------YYQC 354
          ||:  || |:       |||     : :|    |: :    :|  |: :|      :|:
Db    292 DGDFIPADPINLYANAADIDYIAGTNNMDGHIFASID----MPAINKGNKKVTEEDFYKL 347

Qy    355 VHGNLVERRC-GAGTHFSFELQQCDHIELVGCTLPGGESEE---VDVDEDACTGWYCPTE 410
          |    : :   || | |    :             |  |:::   || : |    :  |||
Db    348 VSEFTITKGLRGAKTTFDVYTES-------WAQDPSQENKKKTVVDFETDVL--FLVPTE 398

Qy    411 PIEWEPLPNGCPADFSIDHLLPHES-----------DCGQYLQCVHGQTIARPCPGNLHF 459
          :   |   | :  :|  | |           |    :| | |:  | |      :
Db    399 IALAQHRANAKSAK-TYAYLFSHPSRMPVYPKWVGADHADDIQYVFGKPFATP----TGY 453

Qy    460 SPATQSCESPVTAGCQVFECDSDNQCTSTAAPT-------------------------- 493
           |  ::    : |    |    |     :| ||
Db    454 RPQDRTVSKAMIAYWTNFAKTGDPNMGDSAVPTHWEPYTTENSGYLEITKKMGSSSMKRS 513

Qy    493 --------------AAPTAAPTAAPTAAPTA-APSTVVPPATPPATAAPVPPTTAIPTPA 537
                         | ||     |    ||  : :| ||| |  :  ||||||
Db    514 LRTNFLRYWTLTYLALPTVTDQEATPVPPTGDSEATPVPP-TGDSETAPVPPT------G 566

Qy    538 PTAAPTAAPTTAAPESPTTVTVPPT---AAPTAAPT--TAVPEIPITVTSAPTAAPTAAP 592
           : ||   ||  :   |    ||||    ||   ||  :  | :| |  |     |
Db    567 DSGAPPVPPTGDSGAPP----VPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGD 622

Qy    593 TAAPTAAPT---TAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPT---AAPTTAAPAPN 646
          : ||   ||    | |  ||    ||   ||  : ||     ||||    ||       :
Db    623 SGAPPVPPTGDSGAPPVPPTGDAGPPPVPPTGDSGAP----PVPPTGDSGAPPVTPTGDS 678

Qy    647 TTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLP 695
           |  ||||   : ||| |    :: ||||         | ||  :  :|
Db    679 ETAPVPPT-GDSGAPP-VPPTGDSEAAPV--------PPTDDSKEAQMP 717


Search completed: July 18, 1999, 07:57:21
Job time: 442 sec
```

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 00:41:47 ; Search time 26.5 Seconds
(without alignments)
601.417 Million cell updates/sec

Title: US-09-294-663-3
Perfect score: 4328
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 788

Scoring table: BLOSUM62

Searched: 162890 seqs, 20225328 residues

Database : A_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	520.5	12.0	1404	1	R26049	MSF precursor. New
2	445.5	10.3	1721	1	W48299	Cryptosporidium pa
3	372.5	8.6	763	1	W31852	Mycobacterium tube
4	353.5	8.2	1664	1	W43106	C. thermocellum Ol
5	331.5	7.7	400	1	P91941	Sequence of prepro
6	319.5	7.4	844	1	P60570	Sequence of the Fa
7	318	7.3	572	1	W31855	Mycobacterium tube
8	317	7.3	297	1	W15761	Cotton fibrous tis
9	313	7.2	907	1	P50073	Epstein-Barr virus
10	313	7.2	907	1	R80144	EBV gp350/220. New
11	307.5	7.1	451	1	R92803	Hepatitis A virus
12	292	6.7	278	1	R07671	Intestinal mucin d
13	285.5	6.6	267	1	W64359	Mycobacterium tube
14	285.5	6.6	267	1	W81726	M. tuberculosis im
15	284.5	6.6	278	1	R12535	SMUC-41 intestinal
16	270	6.2	166	1	R07670	Intestinal mucin d
17	268	6.2	166	1	R12534	SMUC-40 intestinal
18	265.5	6.1	214	1	R86913	Cotton fiber-speci
19	245.5	5.7	575	1	R11862	Delta-endotoxin Pv
20	243	5.6	452	1	R80041	Human megakaryocyt
21	238	5.5	849	1	W06725	FLO1 protein, invo
22	237	5.5	486	1	R08034	Unique sequence fr
23	235.5	5.4	1537	1	R60562	Yeast 4.7 kB agglu
24	230.5	5.3	3119	1	W72204	HSV-2 strain SB5 C
25	229	5.3	334	1	W38336	Human kidney injur
26	228.5	5.3	733	1	W69430	Human secreted pro
27	227	5.2	841	1	W34985	Archaebacterial th
28	221	5.1	334	1	P82971	Bioadhesive precur
29	219.5	5.1	629	1	R88466	Drosophila scaveng
30	218	5.0	742	1	R14311	Human milk bile-sa
31	218	5.0	744	1	R45189	BSSL/CEL. DNA enco
32	218	5.0	745	1	R58980	Bile salt-stimulat
33	218	5.0	722	1	R58984	Bile salt-stimulat
34	218	5.0	722	1	R75107	Bile salt-activate
35	218	5.0	722	1	R98926	Human bile salt-ac
36	218	5.0	742	1	R98927	Human bile salt-ac
37	218	5.0	745	1	R99257	Human bile salt-st
38	218	5.0	722	1	W10050	Human bile salt-ac
39	217.5	5.0	132	1	R07673	Intestinal mucin d
40	217.5	5.0	132	1	R12537	SMUC-53 intestinal
41	217.5	5.0	357	1	R66810	Extracellular doma
42	216	5.0	334	1	P83194	Sequence of a bioa
43	216	5.0	2035	1	R57141	Host cell factor p
44	215	5.0	745	1	R20098	Bile-salt stimulat
45	213.5	4.9	918	1	P60723	Sequence of extrac

ALIGNMENTS

RESULT 1
R26049
ID R26049 standard; Protein; 1404 AA.
AC R26049;
DT 02-FEB-1993 (first entry)
DE MSF precursor.
KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
KW stability; proteolytic cleavage; adhesion; alternative splicing.
OS Synthetic.
FH Key Location/Qualifiers
FT region 1. .26
FT /label= Exon_I
FT region 26. .67
FT /label= Exon_II
FT region 67. .107
FT /label= Exon_III
FT region 107. .157
FT /label= Exon_IV
FT region 157. .200
FT /label= Exon_V
FT region 200. .1141
FT /label= Exon_VI
FT region 1411. .1166
FT /label= Exon_VII
FT region 1166. .1212
FT /label= Exon_VIII
FT region 1213. .1266
FT /label= Exon_IX
FT region 1266. .1331
FT /label= Exon_X
FT region 1331. .1373
FT /label= Exon_XI
FT region 1373. .1404
FT /label= Exon_XII
PN WO9213075-A.
PD 06-AUG-1992.
PF 17-JAN-1992; U00433.
PR 18-JAN-1991; US-643502.
PR 10-SEP-1991; US-757022.
PA (GEMY) GENETICS INST INC.
PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
DR WPI; 92-284660/34.
DR N-PSDB; Q27223.
PT New human mega-karyocyte stimulating factors - for treating
PT immune deficiencies, cancer, exposure to radiation or drugs,
PT bacterial and viral infections, etc.
PS Claim 1, 2 and 3; Fig 1; 87pp; English.
CC The sequence given is a full length translation from the megakaryocyte
CC stimulating factor (MSF) precursor. The sequence covered by exons II,
CC III and IV encodes megakaryocyte stimulating factor (MSF). This
CC sequence is modified by the addition of an N-terminal sequence encoding
CC a secretory leader, an initiating methionine proceeding exon II and a
CC terminating codon following exon IV. The cDNA sequence given contains
CC sequences derived from human megakaryocyte colony stimulating factor
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
CC classical mammalian protein secretion signal sequence. The sequence
CC encoding the original meg-CSF includes exons II-IV and is thought to
CC terminate in the region between amino acid residues 134 - 147. The
CC primary transcript of this gene may be cleaved in different ways to
CC yield a family of mRNA's each encoding a different MSF protein. Exons
CC V and VI are thought to be related to the activity of the factor and
CC are also implicated in the stability, folding and processing of the
CC molecule. These exons are also thought to play a role in the observed
CC synergy of MSF with other cytokines. Exons V - XII are believed to be
CC implicated in the processing or folding of the appropriate structure of
CC the resulting factor, ie. one or more of these exons may contain

CC sequences which direct proteolytic cleavage, adhesion, organisation of
CC the cellular matrix or extracellular matrix processing. Both naturally
CC occuring and non-naturally occuring MSF's may be characterised by
CC various combinations of alternatively spliced exons from this sequence,
CC with the exons spliced together in differing orders to form different
CC members of the MSF family.
SQ Sequence 1404 AA;

Query Match 12.0%; Score 520.5; DB 1; Length 1404;
Best Local Similarity 29.5%; Pred. No. 6.1e-21;
Matches 203; Conservative 26; Mismatches 205; Indels 253; Gaps 32;

```
Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
Db     331 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 390

Qy     137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
Db     391 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 450

Qy     186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
Db     451 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 508

Qy     244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
Db     508 ----------------------------------------------PTTTKEPSPTTPK- 521

Qy     304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
Db     521 ----EPAPTTT---------------------------------------------------- 528

Qy     364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
Db     528 --------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 560

Qy     423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
Db     561 APTT-----PKEP----------------------APTTPKKPAPTT----------- 581

Qy     483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
Db     581 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 637

Qy     536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
Db     638 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 695

Qy     590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT-----------VTVPP 634
Db     696 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 755

Qy     635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH-------------APNTTAAPVTTTS 679
Db     756 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 815

Qy     680 ---APATTPEDDDIDP--PLPNDPINP 701
Db     816 DKPAPTTPKETAPTTPKEPAPTTPKKP 842
```

RESULT 2
W48299
ID W48299 standard; Protein; 1721 AA.
AC W48299;
DT 17-AUG-1998 (first entry)
DE Cryptosporidium parvum GP900 antigen.
KW Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW antibody; prophylaxis; treatment; inhibition; retardation;
KW detection; diagnosis; human; 3' region.
OS Cryptosporidium parvum.
FH Key Location/Qualifiers
FT Domain 175. .423
FT /note= "NINC domain"
PN WO9806430-A1.
PD 19-FEB-1998.
PF 11-AUG-1997; U14104.
PR 14-AUG-1996; US-700651.
PA (REGC) UNIV CALIFORNIA.
PI Gut J, Leech J, Nelson RC, Petersen C;
DR WPI; 98-159290/14.
DR N-PSDB; V20700, V20701.
PT Anti-Cryptosporidium antibody - used to develop products for
PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT infections
PS Claim 27; Pages 65-68; 89pp; English.
CC The sequence is that of the GP900 antigen which may be used
CC in the production of anti-Cryptosporidium antibodies. These can be
CC used for the prophylaxis, treatment, inhibition or retardation of
CC a Cryptosporidium infection in humans or in animals such as calves.
CC They can also be used for the detection and diagnosis of related
CC infections.
SQ Sequence 1721 AA;

Query Match 10.3%; Score 445.5; DB 1; Length 1721;
Best Local Similarity 26.0%; Pred. No. 8.2e-17;
Matches 164; Conservative 41; Mismatches 243; Indels 183; Gaps 17;

```
Qy      99 TTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTT 158
Db     270 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTT 329

Qy     159 QAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATT-PAATTPAATT 217
Db     330 TTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKPTTTTTTTTT 389

Qy     218 PAATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYT 277
Db     390 TTTKKPTTTTTATTTTTTS-----------------ETESVIKPDEWCWL--------- 423

Qy     278 FEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFE 337
Db     423 --------------------EKNGECEAKGATY------------VGVIGKDGRIENGM 449

Qy     338 IDWLLPHGNRCDKYYQCVHGNLVERRCGAGTH--FSFELQQCDHIELVGCTLPGGESEEV 395
Db     450 AFTMIPNDD--------------------THVRFRFKVKDVGNTISVRCRKGAGKLEFP 488

Qy     396 DVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPG 455
Db     489 DRS-----------------------LDFTIPPVAGHNS-C-SIIVGVSGD-------G 515

Qy     456 NLHFSPATQSCESPVTAGCQVFECDSDNQC-TSTAAPTAAPTAAPTAAPTAAPSTV 514
Db     516 KIHVSPYGSKDVSLISAPIQPSELFNEVYCDTCTAKYGAIHSGYQTSADFVTTTTAKPTT 575

Qy     515 VPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVP 574
Db     576 T------TTGAPGQPTTT-------------TTGSPSKPTTTT---TTKATTTTTTLNP 612

Qy     575 EIPITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPP 634
Db     613 IITTTTQKPTTTTTTKVPGKPPIATTTTTLKPIVTTTTTKATT--TTTTTVPTTTTTTKR 670

Qy     635 TAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPED------- 688
Db     671 DEMTTTTTPLPD-----------------------IGDIEITPIPIEKMLDKYTRMIY 705

Qy     688 --------DDIDPPLPNDPINPCVEECNVLP 710
Db     706 DYNSGLLLDSNDEPIPGSQAGQIADTSNLFP 736
```

RESULT 3

```
W31852
ID   W31852 standard; Protein; 763 AA.
AC   W31852;
DT   27-APR-1998  (first entry)
DE   Mycobacterium tuberculosis 74 kDa protein.
KW   Tuberculosis; mycobacteria; infection; diagnosis;
KW   antimycobacterial; antibiotic; vaccine.
OS   Mycobacterium tuberculosis.
PN   WO9741252-A2.
PD   06-NOV-1997.
PF   18-APR-1997; E01973.
PR   29-APR-1996; DE-017184.
PA   (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
PI   Espitia C, Honisch C, Moreno C, Singh M;
DR   WPI; 97-549750/50.
DR   N-PSDB; T93610.
PT   New DNA and related proteins or RNA derived from M. tuberculosis -
PT   used for diagnosis of mycobacterial infections, monitoring
PT   vaccination and development of anti-mycobacterial agents
PS   Claim 5; Fig 13; 55pp; English.
CC   This novel 74 kDa protein is encoded by an open reading frame of
CC   a Mycobacterium tuberulosis DNA fragment (see T93610) containing
CC   polymorphic GC-rich sequences.  Its amino acid sequence shows
CC   a high proline content, but there is no homology to any known
CC   proline-rich antigens of mycobacteria.  Novel M. tuberculosis
CC   proteins (see W31851-57) are claimed.  These can be produced as
CC   recombinant proteins, especially in bacterial, yeast, fungal or
CC   higher eukaryote host cells, and used for diagnosing tuberculosis
CC   and other mycobacterial infections in humans or animals.  The
CC   claimed proteins can also be used for epidemiological studies, for
CC   monitoring vaccination, and for the development of vaccines and
CC   anti-mycobacterial drugs.
SQ   Sequence  763 AA;

  Query Match             8.6%;  Score 372.5;  DB 1;  Length 763;
  Best Local Similarity   23.1%;  Pred. No. 3.1e-13;
  Matches  191;  Conservative   40;  Mismatches  290;  Indels  307;  Gaps   35;

Qy      14 VAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYGLKFIAPRDCAPGT 73
           :|  |    |   |    | |  |||   |:|                    |   | |
Db      10 LAPLPPAPPAPAEPKSKPPFPPAPPAPPCWML----------------VSAAPPCPPAP 52

Qy      74 EFKFSAQTCVHAALAGCTLPGPPAETTQA---------PATTQAPTTTQAPTTTTQAPTT 124
                          | ||   ::|         ||   ||    ||    :||
Db      53 -------------------PAPPKPKSKAPFPPVPPAPPARELAPPLPPAP---PEAPRE 90

Qy     125 TTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTT 184
           :  |       |     |    ||    || :   |           ||    ||
Db      91 SRPALPPCPPPPVVIPDPPEPAAPPVPPAPNSPPFPP---------FPPAPKFVPAPPVP 141

Qy     185 QAPTTTQAPTTITQAATTPAATTPAATT-----PAATTPAATTPAATTPGVP-APTSAPV 238
           | :   |     |   ||   |          ||  ||| | |||  | || || | |
Db     142 PVPNSPPFPPFPPAALNPPAPPAPPLANSPPLPPAPPTPAGTPPAAPWPPVPAAPKSKPA 201

Qy     239 WPPICELLP-NGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDS 297
            ||     |    | :|     :| |                :  |       |
Db     202 SPPRPPAPPMPATPMEFPPLPPVPPD------------PISKETPPAPPAPPI-----P 243

Qy     298 PANV---------ECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRC 348
           || |            :| ||||         : :  :|   ||         ||
Db     244 PAPVPIPPVPPLPPVPNKIPPAPPA------PPVAVAAVLVAPCPP-------LP----- 286

Qy     349 DKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP 408
                                                 ||                  |
Db     286 ------------------------------------PLPNNHPPAPPA---------AP 299

Qy     409 TEPIEWEPLPNG---CPADFSIDHL----LPHESDCGQYLQCVHGQTIA---------- 451
               :   ||||     |    :  :    ||          : |: ::
Db     300 VPGVPLAPLPNSHPPAPPSAPVPGVPLAPLP----------ISGRPVSVWKGSFTTLST 348

Qy     451 ---RPCPGNL---HFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAP--------- 496
                 | | |  :     :|:  |  ||:|              |: | |   |

Db     349 FCCRVCSGEVLAGALNPSRPS-RSPLTT------------TTPALPAPIPPLPPLPPLP 394

Qy     496 --TAAP--------TAAPTAAPTAAP---STVVPPA----------TPPATAAP------ 527
             || |         ||      |  ||   |  ||||          |||  ||
Db     395 INTAVPPIPPLPPVTALAPPLPPLAPLPISPGVPPAPPIPPGKPWTTPPLAPAPPEPKTV 454

Qy     527 --VPPTTAIP------------TPAPTAAPTAAPTTAAPESPTTVTVPPT----AAPTAA 568
             :||  : |              | | ::|    |   ||  |: | |||:     || ||
Db     455 PVLPPGPSCPPSEKPNPPAPPEPPEPKSSPALPPAPPAPSMPSAVRVPPSPPIPPAPPAA 514

Qy     569 PTTAVPEIPITVTSAPTA----------------APTAAPT------AAPTAAPTTAVP 605
           |  ::| :|    | |                  || | ||      | |   |    |
Db     515 PRASMPALPPAPPSPPATRLCPPLPPSPPAPNSPPAPPAPPTPPKLLSANPPCPPVPPAP 574

Qy     606 EIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVA 665
           |    :||      | | | |   || |    | |||      ||:| |   ||
Db     575 NRPPAPPAPPAPPELPAPPDPPT----PPVANSPPAPPAP----PAPPSALPFVNPP--- 624

Qy     666 HAPNTTAAPVTTTSAP------------ATTPEDDDIDPPLPNDPINP 701
           || | |||  :  : |            ||||      || ||    |
Db     624 -APPTPAAPKSRPALPAAPPAPPAPPVRATTPPPAPPAPPAPNSMALP 670


RESULT   4
W43106
ID   W43106 standard; Protein; 1664 AA.
AC   W43106;
DT   16-OCT-1998  (first entry)
DE   C. thermocellum OlpB protein.
KW   Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
KW   cohesin domain; catalytic subunit; scaffold subunit; SdbA; synergistic;
KW   cellulosome integrating protein; scaffoldin dockerin binding protein.
OS   Clostridium thermocellum.
FH   Key             Location/Qualifiers
FT   Domain          28. .192
FT                   /note= "cohesin type II domain"
FT   Domain          207. .363
FT                   /note= "cohesin type II domain"
FT   Domain          409. .565
FT                   /note= "cohesin type II domain"
FT   Domain          607. .763
FT                   /note= "cohesin type II domain"
PN   FR2748479-A1.
PD   14-NOV-1997.
PF   10-MAY-1996; 005854.
PR   10-MAY-1996; FR-005854.
PA   (INSP ) INST PASTEUR.
PI   Beguin P, Leibovitz E;
DR   WPI; 98-011569/02.
DR   N-PSDB; T86623.
PT   Cellulase proteins with cohesin or dockerin type II domains - useful
PT   for potentiating the activity of multiprotein enzyme complexes
PS   Claim 7; Page 31-39; 60pp; French.
CC   Multimeric protein, especially enzymatic, complexes are held together
CC   by protein-protein interactions between domains designated dockerins
CC   and cohesins, which are found on the catalytic and scaffold subunits
CC   respectively.  An example of such a complex is the cellulose degrading
CC   protein complex from Clostridium thermocellum, known as the cellulosome.
CC   This complex comprises around 15 proteins including endoglucanases,
CC   cellobiohydratases, hemicellulases, e.g. xylanases or lichinases, which
CC   interact with a central "scaffold" protein designated the cellulosome
CC   integrating protein (CipA; see W43108).  The catalytic subunits interact
CC   with the Cip subunit via conserved 23 amino acid dockerin domains. Cip
CC   has been shown to contain 9 copies of a cohesin domain.
CC   The invention relates to the isolation of proteins binding to a novel
CC   dockerin type domain found in the C-terminal portion of Cip.  The new
CC   domain is designated a type II dockerin domain (as compared to the type I
CC   domain found on the catalytic subunits of the cellulosome).  The type II
CC   dockerin domain has some sequence similarity to the type I dockerins but
CC   is unable to bind type I cohesin domains.
CC   The sequence presented here is an example of a protein which binds
CC   the novel type II dockerin domain and is the product of the OlpB gene.
CC   The protein contains 4 type II cohesin domains in the N-terminal portion
```

CC of which the first domain (amino acid residues 28-192) is thought to bind
CC CipA. The novel type II dockerin and cohesin domains can be used in
CC complexes, especially enzyme complexes, to potentiate their catalytic
CC actions in a synergistic manner.
SQ Sequence 1664 AA;

Query Match 8.2%; Score 353.5; DB 1; Length 1664;
Best Local Similarity 23.5%; Pred. No. 7.2e-12;
Matches 149; Conservative 70; Mismatches 277; Indels 139; Gaps 23;

```
Qy      93 PGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 152
           | | | | :   | :   | : |     || |     | :  || :  || :  || :   .
Db     917 PTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSD 976

Qy     153 APTTTTQ-------APTTTTQAP---TTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATT 202
           || : :         |: | : |    | :  || :  || :  || :  ||   :   :
Db     977 EPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1036

Qy     203 PAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPH 262
              |     || : ||    |  |    | |:  |   |  |  |:  |   |
Db    1037 DEPTPSDEPTP-SETPEEPIPTDTPSDEPTPSDEPT--PSDEPTPSDEPTPSD------- 1087

Qy     263 DKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDI 322
                 : :   | |:  |           |:|::     | :|:   |  :|
Db    1087 --------EPTPSETPEEPIP----------TDTPSD-----EPTPSDEPTPSDEPTPS 1122

Qy     323 DIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIEL 382
           |     |   |:    :  :|     |                                |
Db    1123 DEPTPSDEPTPSETP-EEPIPTDTPSD------------------------------EP 1150

Qy     383 VGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQ 442
                |    |    ||   :    |:|  | ||:|   |:|       |  ||
Db    1151 TPSDEPTPSDEPTPSDEPTPSDEPTPSETPE-EPIPTDTPSDEPTPSDEPTPSD------ 1204

Qy     443 CVHGQTIARPCPGNLHFSPA-----TQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTA 497
                 | | :   :|:     ::: | |:       |    :: | :  ||  |:
Db    1204 --------EPTPSD-EPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPT--PSD 1252

Qy     498 APTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTV 557
            || :   |:  |   :|  ||     |    |    | |:  || :      : ||
Db    1253 EPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT-- 1311

Qy     558 TVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSP-PT 616
              |:  || : |   || ||     ||  |:  ||  |:  || :  : ||    | |:
Db    1311 ---PSDEPTPSET---PEEPI-----PTDTPSDEPT--PSDEPTPS--DEPTPSDEPTPS 1355

Qy     617 AAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAP--------------P 662
              ||   |:   | :  || : |   | | || |  |:  ||:
Db    1356 DEPT---PSDEPTPSDEPTPSETPEEPTPTTTPTPTPSTTPTSGSGGSGGSGGGGGGGGG 1412

Qy     663 TVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPND 697
           ||  :|       | || | :||   :|: | |:|
Db    1413 TVPTSPTP-----TPTSKPTSTPAPTEIEEPTPSD 1442
```

RESULT 5
P91941
ID P91941 standard; peptide; 400 AA.
AC P91941;
DT 07-FEB-1990 (first entry)
DE Sequence of preprospasmolysin.
KW Preprospasmolysin; Xenopus laevis.
OS Xenopus laevis
PN DE3808456-A.
PD 28-SEP-1989.
PF 14-MAR-1988; 808456.
PR 14-MAR-1988; DE-808456.
PA (PLAC) Max-Plank Ges. Wissensch.
PI Weickmann H, Fincke K, Weickmann F, Huber B, Liska H, Prechtel J;
DR WPI; 89-286097/40.
DR N-PSDB; N91235
PT Polypeptide preprospasmolysin and its fragments - derived from Xenopus
PT laevis, with spasmolytic and neurotransmitter activities.
PS Claim 1; page 5; 8pp; german.
CC The peptide has spasmolytic and neurotransmitter functions. Fragments
CC comprising residues 21-400, 21-69, 351-400 and 72-348 may also be
CC produced. One or more of residues 1,4,5,8,9,11,20,21,22,24,25,26,33,35,
CC 36 and 42-9 may be replaced by another amino acid present in the same
CC position in one of the other fragments, and/or contg. the exchange D-E,
CC I-L, K-R, N-Q and/or S-T.
SQ Sequence 400 AA;

Query Match 7.7%; Score 331.5; DB 1; Length 400;
Best Local Similarity 29.9%; Pred. No. 2.7e-11;
Matches 115; Conservative 21; Mismatches 142; Indels 107; Gaps 21;

```
Qy     434 ESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAG-CQVFECDSD---------- 483
           |:|| : : |    :|             |: |    |||  :  ||  |
Db      40 EADC-RAVGCCFDSSILN----------TKWCFYNATAGPIKKLECSGDPTKRIDCGFP 87

Qy     483 ----NQC-----------------TSTAAPTAAP---TAAPTAAPTAAPTAAPSTVVPP 517
               ||                    |   | ||   ||: ||  |  ||  | |   |
Db      88 RITEKQCILRGCCFDSSISGVKWCYARTVITTPAPDTTTASTTAETTTVPT-TPETTTVP 146

Qy     518 ATPPATAAP-VPPTTAIP-TPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPE 575
           ||  |  |  | || :| || |  ||   ||  | :| | ||| |   |  ||| ||
Db     147 TTPETTTVPTTPETTTVPTTPETTTVPTTPETTTVPTTPETTTVPTT--PE 204

Qy     576 IPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPT 635
               |: ||   |   |  ||   || || .| | |:  ||  ||    | |  |  ||
Db     205 ----TTTVPTTPET---TTVPTTPETTTVPTTPETTTASTTAETTTVPTTPET--TTEPT 255

Qy     636 AAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLP 695
             |||          | |||  ||  | |        |    || :   || |     ||
Db     256 TTPTTD--------TTPPTLPPT--PETTTETTTETTTETTTETTTETTTAPPP-- 304

Qy     696 NDPINPCVE---ECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFN---PTTKTCDF-- 748
               |     :|       |||:                :|  |:   | || | :
Db     304 -----ECAADRVDCGYSGITQADCE--------------GKGCIFDSTIPETKWCFYTE 343

Qy     748 ----ACNVGC-VRSNIQMSESYEGV 767
               |    | |  :::    | |:
Db     344 AEAPARKAECTVDPSVRTDCGYPGI 368
```

RESULT 6
P60570
ID P60570 standard; Protein; 844 AA.
AC P60570;
DT 24-AUG-1991 (first entry)
DE Sequence of the Falciparum Interspersed Repeat Antigen
DE (FIRA).
KW Malaria vaccine; antigen; epitope.
OS Plasmodium falciparum.
PN WO8601802-A.
PD 27-MAR-1986.
PF 11-SEP-1985; 006960.
PR 11-SEP-1984; AU-007067.
PR 11-SEP-1984; AU-007066.
PR 10-SEP-1985; AU-047326.
PA (HALL-) HALL INST MED RES.
PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
DR WPI; 86-094065/14.
DR N-PSDB; N60473.
PT DNA coding for Plasmodium falciparum antigens - expressing
PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P
PT falciparum
PS Disclosure; Fig 7; 55pp; English.
CC The inventors claim a novel DNA molecule which comprises a
CC nucleotide sequence corresp. to all or a portion of the base
CC sequence coding RESA (N60472) or FIRA (N60473). RESA and FIRA have
CC antigenicity suitable for providing protective immunity against
CC Plasmodium falciparum malarial infections.
SQ Sequence 844 AA;

```
  Query Match             7.4%;  Score 319.5;  DB 1;  Length 844;
  Best Local Similarity   22.0%;  Pred. No. 2.5e-10;
  Matches  179;  Conservative   28;  Mismatches  229;  Indels  379;  Gaps   33;

Qy      96 PAET--TQAPATTQAPTTTQAPTTTTQAPTT-----TTQAPTT-----TTQAPTTTQAPT 143
           | |||  || | | : | ||| |  ||| | |     | | | |     | | | | | |
Db     164 PQETVETQEPVTIEEPVTTQEP-VTTQEPITVQEPVTVQEPVTVQEPVTVQEPVTVQEPV 222

Qy     144 TTQAPTTTQAPTT-----TTQAPTT-----TTQAPTT-----TTQAPTTTQAPTTTQAPT 188
           | | | | | | |     | | | |     | | | |     ||| | ||| | | : |
Db     223 TVQEPVTVQEPVTSQEPVTPQEPVTPQEPVTPQEPVTPQEPVTTQEPVTTQEPVTIEEPV 282

Qy     189 TTQAPTTITQAATTPAATT---PAAT-TPAATTPAATTPAATTPGVPAPTSAPVWPPICE 244
           ||| | || :  ||    |   | | |  |    ||    |   |  | ||
Db     283 TTQEPVTIEEPVTTQEPVTIEEPVTTQEPVTTQEPVTTQEPVTTQEPV------ 337

Qy     245 LLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECD 304
                                       | |:   |
Db     337 ------------------------------TVEEHIDE------------------- 345

Qy     305 GEISPAPPVTEGNEDEDIDIGDLL----DNGCPANFEIDWL-LPHGNRCDKYYQCVHGNL 359
                  :|:| ::| :  |     :       :  ||    ||: ||       :
Db     345 ---------KKGSEGDNISLSSLSEETEEKSHTKKKKSSWLKFGRGNKNDK------KSK 389

Qy     360 VERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPN 419
           |::                    |  || : : ||
Db     390 NEKK----------------------PSLESVKQNADEQ------------------K 407

Qy     420 GCPADFSIDHLLPHESDCGQYLQCVHGQ---TIARPCPGNLHFSPATQSCESPVTAGCQV 476
           | ||  |              |: |   ||  |         |||  | | |
Db     408 EQPTDSQIS--------------VNAQDSVTIQEP--------TATQ--EPPTT----- 438

Qy     477 FECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPAT--PPATA-APVPPTTAI 533
                : |:|  ||          ||   |    |   | |   | ||  ||     |
Db     438 ------QELTATQEPTTTQETVTEQEPTT--TQETVTAQEPITTQEPVTAQEPVTTQELI 489

Qy     534 PTPAPTAAPTAAPTTAA---------------------------------------------- 551
           |  |:     |    |
Db     490 ATQEPSTTQEHADEKKASEGDNISLSRLSEETEEKSHTKKKKSSWLKFGRGNKNDKKSKN 549

Qy     551 ----------------PESPT----------TVTVPPTAAPTAAPTTAVPEIPITVTSAP 584
                           | ||           :||:    |    |||      :| | |
Db     550 EKKPSLESVKQNADEQKEQPTDSQISVNAQDSVTIQEPTATQEPPTTQ----ELTATQEP 605

Qy     585 TAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTA-APTT-----AAPAPNTT--------- 630
           |      ||    |   : | |   | ||  | |      |   |:||
Db     606 TTTQETVTEQEPTTTQETVTAQEPITTQEPVTAQEPVTTQELIATQEPSTTQEHADEKKA 665

Qy     630 ------------------------------------------------------------ 630

Db     666 SEGDNISLSRLSEETEEKSHTKKKKSSWLKFGRGNKNDKKSKNEKKPSLESVKQNADEQK 725

Qy     630 ---------------VTV--------PPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTV-A 665
                           ||:        |||    ||   | ||         ||    || |
Db     726 EQPTDSQISVNAQDSVTIQEPTATQEPPTTQELTATQEPTTTQETVTEQEPTTTQETVTA 785

Qy     666 HAPNTTAAPVT-----------TTSAPATTPEDDD 689
             | ||  |||            |  |:|| |  |
Db     786 QEPITTQEPVTAQEPVTTQELIATQEPSTTQEHAD 820


RESULT   7
W31855
ID   W31855 standard; Protein; 572 AA.
AC   W31855;
DT   27-APR-1998  (first entry)
DE   Mycobacterium tuberculosis 55 kDa protein.
KW   Tuberculosis; mycobacteria; infection; diagnosis;
KW   antimycobacterial; antibiotic; vaccine.
OS   Mycobacterium tuberculosis.
PN   WO9741252-A2.
PD   06-NOV-1997.
PF   18-APR-1997; E01973.
PR   29-APR-1996; DE-017184.
PA   (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
PI   Espitia C, Honisch C, Moreno C, Singh M;
DR   WPI; 97-549750/50.
DR   N-PSDB; T93610.
PT   New DNA and related proteins or RNA derived from M. tuberculosis -
PT   used for diagnosis of mycobacterial infections, monitoring
PT   vaccination and development of anti-mycobacterial agents
PS   Claim 11; Fig 16; 55pp; English.
CC   This novel 55 kDa protein is encoded by an open reading frame of
CC   a Mycobacterium tuberculosis DNA fragment (see T93610) containing
CC   polymorphic GC-rich sequences.  Its amino acid sequence shows
CC   a high proline content, but there is no homology to any known
CC   proline-rich antigens of mycobacteria.  Novel M. tuberculosis
CC   proteins (see W31851-57) are claimed.  These can be produced as
CC   recombinant proteins, especially in bacterial, yeast, fungal or
CC   higher eukaryote host cells, and used for diagnosing tuberculosis
CC   and other mycobacterial infections in humans or animals.  The
CC   claimed proteins can also be used for epidemiological studies, for
CC   monitoring vaccination, and for the development of vaccines and
CC   anti-mycobacterial drugs.
SQ   Sequence   572 AA;

  Query Match             7.3%;  Score 318;  DB 1;  Length 572;
  Best Local Similarity   22.4%;  Pred. No. 2e-10;
  Matches  167;  Conservative   41;  Mismatches  272;  Indels  264;  Gaps   33;

Qy      20 VSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSA 79
           |  | |:     | |  ||                                  | |  :|
Db       1 VPAAPKSKPASPPRPPAPP------------------------------MPATPMEFPP 29

Qy      80 QTCVHA-ALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTQAPTTTQAPTT 138
              |    ::  | | |||     ||    |     |    : |      |
Db      30 LPPVPPDPISKETPPAPPAPPI-PPAPVPIPPVPPLPPVPNKIPPAPPAPPVAVA---AV 85

Qy     139 TQAPTTTQAPTTTQAPTTTTQAP---TTTTQAPTTTTQAPTTTQAPTTTQAPTTTQA-PT 194
             ||      |     |    ||         | :   || :   |    ||      |
Db      86 LVAPCPPLPPLPNNHPPAPPAAPVPGVPLAPLPNSHPPAPPSAPVPGVPLAPLPISGRPV 145

Qy     195 TITQAATTPAAT-----------------TPAATTPAATTPAATTPAATTPGVP----- 232
           :: : : |  :|                 |: :    ||||    |    | :|
Db     146 SVWKGSFTTLSTFCCRVCSGEVLAGALNPSRPSRSPLTTTTPALPAPIPPLPPLPPLPIN 205

Qy     232 -APTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNP 290
            |    |  ||:  | |  |        | |           | |       | |    |
Db     206 TAVPPIPPLPPVTALAPPLPP-------LAP--------LPISPGVPPAPPIPPG---KP 247

Qy     291 YVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDK 350
           :    :|        ::||||     | : : :                  || |  |
Db     248 WT----TP-------PLAPAPP-----EPKTVPV----------------LPPGPSC-- 273

Qy     351 YYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTE 410
                                                 |  |                | |
Db     273 -----------------------------------PPSEKPNPP----------APPE 285

Qy     411 PIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPV 470
           |    | |   ||       ||                    | |       | :   ||
Db     286 P----PEPKSSPA-------LP------------------PAPPAPSMPSAVRVPPSP- 315

Qy     471 TAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPST----VVPPATPPATAAP 526
                              || ||| |:  | |  || : |:|     :||: |   : |
Db     315 ---------------PIPPAPPAAPRASMPALP-PAPPSPPATRLCPPLPPSPPAPNSPP 358

Qy     527 VPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTS---- 583
           ||    |    :| |   |   ||  |     ||  ||   |    |   |  |
Db     359 APPAPPTPPKLLSANPPCPPVPPAPNRPPAPPAPP--APPELPAPPDPPTPPVANSPPAP 416

Qy     583 -APTAAPTAAPTAAPTAAPTTAVPEI-PTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTT 640
           || | |:| |   | | || | |:  |    :||      | |||    | || | |
Db     417 PAPPAPPSALPFVNPPAPPTPAAPKSRPALPAAPP------APPAPPVRATTPPPAPP-- 469
```

```
Qy    641 AAPAPNTTVTVPPT----------AAPTA-----------APPTVAHAPNTTAAPVTTTSA 680
          | |||| :: :||          | | |           |||    ||:  | |:|
Db    469 APPAPN-SMALPPAPPDPPIPLLATPPAPPAPPLPMSPPAPPLPPAAPDPPAPPLTINQP 527

Qy    681 PATTPEDDDIDPPL---PNDPINP 701
          |:         |||   |  |: |
Db    528 PS---------PPLAPVPGAPLAP 542
```

RESULT 8
W15761
ID W15761 standard; Protein; 297 AA.
AC W15761;
DT 19-JUN-1997 (first entry)
DE Cotton fibrous tissue specific protein KC03.
KW Fibrous tissue; cotton fibre; cotton flower; Gossypium barbadense;
KW Gossypium hirsutum.
OS Gossypium barbadense.
PN J09075093-A.
PD 25-MAR-1997.
PF 20-FEB-1996; 031987.
PR 21-FEB-1995; US-391696.
PR 29-DEC-1995; US-580545.
PA (TOYM) TOYOBO KK.
PA (UYTE-) UNIV TEXAS TECH.
DR WPI; 97-239274/22.
DR N-PSDB; T60569.
PT Cotton fibrous tissue gene - used to produce transformants with
PT improved cotton fibres, and improved yield
PS Claim 3; Page 21-22; 37pp; Japanese.
CC W15759-W15763 represent cotton fibrous tissue proteins of the invention.
CC The genes encoding these sequences are specifically expressed during the
CC growth of cotton fibre. The genes were isolated from the polyA RNA of the
CC ovules of a cotton flower of Gossypium barbadense, and G. hirsutum. The
CC genes encoding these sequences, and vectors containing them are used in
CC the preparation of cotton transformants, with improved cotton fibres and
CC improved yield.
SQ Sequence 297 AA;

Query Match 7.3%; Score 317; DB 1; Length 297;
Best Local Similarity 35.8%; Pred. No. 1.2e-10;
Matches 101; Conservative 8; Mismatches 99; Indels 74; Gaps 17;

```
Qy    497 AAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTT 556
          |||   |  ||  |    ||  || |  ||||  | |||||  ||  ||   |  ||
Db     35 AAP--VPVKAPIPA-----PPVKPPTT--PVPPYKA-PTPAP---PTKGPTPYKP--PTK 79

Qy    557 VTVPPTAAPTAAPTTAV-PEIPITVTSAPT--------AAPTAAPT--------AAPTAA 599
            ||   || |||    |  |   | |||        | || |||        | || |
Db     80 APTPPYKPPTKAPTPPYKPPAPAPPTKAPTPPYKPPAPAPPTKAPTPPYKPPAPAPPTKA 139

Qy    600 PTTAVPEIPTTVTSPPTAAPT------TAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPP 653
          ||   |       :||| |||      | ||||      ||   ||    || |   ||
Db    140 PT---PPFKPPAPAPPTKAPTPPYKPPTPAPAPPVKAPTPPYMPPTPPTKAP-TPAPAPP 195

Qy    654 TAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAH 713
          | |||  ||      |    | || : |   |      ||||  |:
Db    196 TKAPT--PPYKPPVPTPPVKPPTTPAPPYKPP-----SPPLP--PVR-----------TK 235

Qy    714 ADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVR 755
           ||            : :| :  : :  |  |  ||   | |
Db    236 KDC-----------IPLCGQRCKLHSRTNLCLRACMTCCDR 265
```

RESULT 9
P50073
ID P50073 standard; Protein; 907 AA.
AC P50073;
DT 17-OCT-1991 (first entry)
DE Epstein-Barr virus (EBV) outer surface protein.
KW Epstein-Barr virus; antigen; vaccine.
OS Epstein-Barr virus.
PN EP-151079-A.
PD 07-AUG-1985.
PF 28-JAN-1985; 400141.
PR 30-JAN-1984; US-575352.
PR 23-JUL-1984; US-633558.
PA (UYCH-) UNIV OF CHICAGO.
PI Kleff E, Tanner J, Hummel M, Belsel C;
DR WPI; 85-191978/32.
DR N-PSDB; N50114.
PT New fragment of Epstein-Barr Virus DNA - useful in vector to
PT express polypeptide for use in prepn. of vaccine against the
PT virus and for use in diagnosis.
PS Claim 2; Page 23-25; 26pp; English.
CC The sequence encodes an outer surface viral protein of EBV, used
CC to generate antibodies reacting with the surface proteins of
CC EBV-infected cells, and in the preparation of a vaccine against EBV.
SQ Sequence 907 AA;

Query Match 7.2%; Score 313; DB 1; Length 907;
Best Local Similarity 22.5%; Pred. No. 6e-10;
Matches 168; Conservative 65; Mismatches 221; Indels 294; Gaps 32;

```
Qy     43 WLLPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGC-TLPGPPAET-- 100
          |  |:  :    : |:: |    |  |      |::       ::|  | |     |
Db    352 WAWPNNTETD--FKCKWTLTSGTPSGC-ENISGAFASNRTFDITVSGLGTAPKTLIITRT 408

Qy    100 -TQAPATTQAPTTTQAPTTTTQAPTTTTQA---PTTTTQAPTTTQAPTTTQAPTTTQAPT 155
           | |  ||     ::|| :|| :||  |     | |||  |::|  ||   || :|  ||
Db    409 ATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPAST-GPT 467

Qy    156 TTTQAPTTTTQAPTTTTQAPTT----------TQAP--TTTQAPTTTQAP--TTITQAA 200
          :|   |: | | ||:  :| |           ::||  |:: :| ||  |  |: | |
Db    468 VSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTSSTSPVTTPTPNATSPTPAV 527

Qy    201 TTPAATTPAAT--TPAATTPAATTPAATTP--GVPAPTSAPVWPPICELLPNGCPADFDI 256
          |||   || ||  ||| |||   || ||:|  |  :|||   |      ||
Db    528 TTP---TPNATSPTPAVTTP---TPNATSPTLGKTSPTSAVTTP-----TPN-------- 569

Qy    257 HLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEG 316
                                                  ||      |: ||   ||
Db    569 ------------------------------------ATSPTL----GKTSPTSAVT-- 585

Qy    317 NEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQ 376
                                          |  |
Db    585 -------------------------TPTPN------------------------------ 590

Qy    377 CDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESD 436
                | |                 |   ||  :   | ||
Db    590 --------ATSP-------------TLGKTSPTSAVT-TPTPNAT-------------- 613

Qy    437 CGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPT 496
                | |:    |                        : :: |      :||   |
Db    613 ---------GPTVGETSP----------------------QANATNHTLGGTSPTPVVT 640

Qy    497 AAPTAAPTAA-----------------PTAAPSTVVPPATPPATAAPVPPTTAIPT--- 536
          : |  | :|                  |:: | |: |  :  :|:     :| ||
Db    641 SQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSDNSTSHMPLLTSAHPTGGE 700

Qy    536 ----PAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVP-EIPITVTSAP-----T 585
              | :  |   :|::|| |   |    : |  : |:  | |: :|  : |
Db    701 NITQVTPASISTHHVSTSSPE-PRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNATSP 759

Qy    586 AAPTAAPTAAPTAAPTTAVPEIPT-----------TVTSPPT---AAPTTAAPAPNTTVT 631
           ||:   || ||   |       |            | | | |       ||  |  | |
Db    760 QAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRPRYNATTY 819

Qy    632 VPP------------TAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVT--- 677
          :||            |: | | | |   || ||||: |  :  ::      : | :|
Db    820 LPPSTSSKLRPRWTFTSPPVTTAQA---TVPVPPTSQPRFSNLSMLVLQWASLAVLTLLL 876

Qy    677 ------------TTSAPATTPEDDDID 691
```

```
                    :||   |||  || :
Db    877 LLVMADCAFRRNLSTSHTYTTPPYDDAE 904


RESULT  10
R80144
ID   R80144 standard; Protein; 907 AA.
AC   R80144;
DT   18-JAN-1996  (first entry)
DE   EBV gp350/220.
KW   EBV; gp350; gp220; gp350/gp220; non-splicing variant; vaccine.
OS   Epstein-Barr virus.
FH   Key             Location/Qualifiers
FT   peptide         1. .18
FT                   /label= Sig_peptide
FT   region          861. .881
FT                   /note= "transmembrane region"
PN   WO9528488-A1.
PD   26-OCT-1995.
PF   13-APR-1995; U04611.
PR   18-APR-1994; US-229291.
PA   (AVIR-) AVIRON.
PI   Jackman WT,  Spaete R;
DR   WPI; 95-373802/48.
DR   N-PSDB; T04821.
PT   New DNA encoding a homogeneous gp350 protein - can be used for
PT   preventing and treating Epstein-Barr virus-related diseases or
PT   conditions
PS   Disclosure; Fig.1; 61pp; English.
CC   The donor and acceptor splice sites of the EBV gene encoding gp350/
CC   220 are mutated by replacement of native nucleotides by non-native
CC   nucleotides, without altering the encoded amino acid sequence,
CC   resulting in elimination of gp220 prodn.  Recombinant homogeneous
CC   gp350, useful in vaccines, is expressed in mammalian or insect cell
CC   hosts.
SQ   Sequence  907 AA;


  Query Match           7.2%;  Score 313;  DB 1;  Length 907;
  Best Local Similarity   22.5%;  Pred. No. 6e-10;
  Matches  168;  Conservative   65;  Mismatches  221;  Indels  294;  Gaps   32;

Qy     43 WLLPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGC-TLPGPPAET-- 100
          |  |:  :    : |:: |    |  |       |::           ::|  | |     |
Db    352 WAWPNNTETD--FKCKWTLTSGTPSGC-ENISGAFASNRTFDITVSGLGTAPKTLIITRT 408

Qy    100 -TQAPATTQAPTTTQAPTTTTQAPTTTTQA---PTTTTQAPTTTQAPTTTQAPTTTQAPT 155
           | |  ||     ::|| :|| :||  |      | |||  |::|  ||   || :|  ||
Db    409 ATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPAST-GPT 467

Qy    156 TTTQAPTTTTQAPTTTTQAPTT-----------TQAP--TTTQAPTTTQAP--TTITQAA 200
           :|   |: | | ||:  :| |           ::||  |:: :| ||  |  |: | |
Db    468 VSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTSSTSPVTTPTPNATSPTPAV 527

Qy    201 TTPAATTPAAT--TPAATTPAATTPAATTP--GVPAPTSAPVWPPICELLPNGCPADFDI 256
          |||   || ||  ||| |||   || ||:|  |  :||||   |      ||
Db    528 TTP---TPNATSPTPAVTTP---TPNATSPTLGKTSPTSAVTTP-----TPN-------- 569

Qy    257 HLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEG 316
                                                 ||      |: ||   ||
Db    569 ------------------------------------ATSPTL----GKTSPTSAVT-- 585

Qy    317 NEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQ 376
                                      |  |
Db    585 ------------------------TPTPN---------------------------- 590

Qy    377 CDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESD 436
                |  |               |   ||  :   | ||
Db    590 --------ATSP-------------TLGKTSPTSAVT-TPTPNAT---------------- 613

Qy    437 CGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPT 496
                | |:   |                        : :: |      :||   |
Db    613 ---------GPTVGETSP----------------------QANATNHTLGGTSPTPVVT 640

Qy    497 AAPTAAPTAA-----------------PTAAPSTVVPPATPPATAAPVPPTTAIPT--- 536
          : |  | :|                    |:: | |: |  :  :|:     |:| ||
Db    641 SQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSDNSTSHMPLLTSAHPTGGE 700

Qy    536 ----PAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVP-EIPITVTSAP-----T 585
                | :  |   :|::|| |   |    : |  : |:  | |: :|  : |
Db    701 NITQVTPASISTHHVSTSSPE-PRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNATSP 759

Qy    586 AAPTAAPTAAPTAAPTTAVPEIPT-----------TVTSPPT---AAPTTAAPAPNTTVT 631
           ||:   || ||   |         |           | | | |      ||  |  | |
Db    760 QAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRPRYNATTY 819

Qy    632 VPP------------TAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVT--- 677
          :||            |: | | | |   || ||||: |  :  ::      : | :|
Db    820 LPPSTSSKLRPRWTFTSPPVTTAQA---TVPVPPTSQPRFSNLSMLVLQWASLAVLTLLL 876

Qy    677 ------------TTSAPATTPEDDDID 691
                       :||   |||  || :
Db    877 LLVMADCAFRRNLSTSHTYTTPPYDDAE 904


RESULT  11
R92803
ID   R92803 standard; Protein; 451 AA.
AC   R92803;
DT   24-MAY-1996  (first entry)
DE   Hepatitis A virus receptor.
KW   Hepatitis A virus; receptor; HAV; diagnosis; therapy; vaccine;
KW   transgenic animal.
OS   Cercopithecus aethiops.
PN   WO9604376-A1.
PD   15-FEB-1996.
PF   04-AUG-1995; U09941.
PR   05-AUG-1994; US-287001.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Feinstone SM,  Kaplan G;
DR   WPI; 96-129392/13.
DR   N-PSDB; T17864.
PT   DNA encoding Hepatitis A virus receptor and its fragments - useful
PT   for the detection, diagnosis, prevention and treatment of HAV
PT   infection
PS   Claim 1; Page 39-41; 52pp; English.
CC   A hepatitis A virus (HAV) receptor (R92803) was identified as the
CC   product of a nucleic acid sequence isolated from HAV-infected
CC   primary African green monkey kidney cells.  Recombinant receptor
CC   can be produced in large amounts by expression of the nucleic acid
CC   in prokaryotic or eukaryotic host cells or in transgenic animals.
CC   It can be used in the detection, prevention and treatment of HAV
CC   infection, in methods of purifying or removing HAV from samples,
CC   and methods of determining the anti-HAV activity of cpds.
CC   Expression in transgenic animals provides a method for testing
CC   vaccine efficacy.
SQ   Sequence  451 AA;


  Query Match           7.1%;  Score 307.5;  DB 1;  Length 451;
  Best Local Similarity   41.6%;  Pred. No. 5.9e-10;
  Matches   87;  Conservative    4;  Mismatches   53;  Indels   65;  Gaps   10;

Qy     94 GPPAET----------TQAPATTQAPTTTQAPTTTTQAPTTTQAPTT----TTQAPTTT 139
          |||  |          |  | ||  |||| |||||  |||||   ||    ||  ||||
Db    125 GPPRVTTPIVRTVRTSTTVPTTTTLPTTTTLPTTTT-LPTTTTLPMTTTLPMTTTHPTTT 183

Qy    140 QAPTTTQAPTTTQAPTT---TTQAPTTTTQAPTTTT-----QAPTTTQAPTTTQAPTTTQ 191
           ||||  ||||  |||   ||  ||| |  |||||       | ||  ||||  ||||
Db    184 TVPTTTTLPTTTTLPTTLPMTTTLPTTRT-LPTTTTLPTTMTLPMTTTLPTTTTLPTTTT 242

Qy    192 APTTITQAATTPAATT-------PAATT-PAATTPAATT-----------------PAAT 226
           |||     | |  ||       |  || |  ||   ||                   | ||
Db    243 LPTTTLPTMTLPTTTTLPTMMTLPTTTTLPTTTTLPTTTMVSTFVPPTPLPMQNHEPVAT 302

Qy    227 TPGVP----------------APTSAPVW 239
```

```
          :|  |                  |||:|::
Db    303 SPSSPQPAETHPVTLLGATRTQPTSSPLY 331


RESULT  12
R07671
ID   R07671 standard; protein; 278 AA.
AC   R07671;
DT   22-FEB-1991  (first entry)
DE   Intestinal mucin deduced from clone SMUC 41.
KW   Epithelial cancer; cystic fibrosis; ulcerative colitis;
KW   glycosyl transferase; familial polyposis coli;
KW   glycosylation patterns.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   region          6. .28
FT                   /label= repeat 1
FT   region          29. .52
FT                   /label= repeat 2
FT   region          53. .75
FT                   /label= repeat 3
FT   region          76. .98
FT                   /label= repeat 4
FT   region          99. .121
FT                   /label= repeat 5
FT   modified_site   265. .267
FT                   /label= N-glycosylation site
PN   WO9012892-A.
PD   01-NOV-1990.
PF   25-JUL-1989; U03206.
PR   14-APR-1989; US-338710.
PA   (REGC) University of California.
PI   Kim YS; Gum JR
DR   WPI; 90-348495/46.
DR   N-PSDB; Q06484.
PT   Human intestinal mucin DNA, polypeptides and antibodies - used in
PT   diagnosis and treatment of e.g. cancers, cystic fibrosis and
PT   colitis.
PS   Disclosure; Fig 4B; 45pp; English.
CC   The sequence was deduced from a clone isolated from a cDNA library
CC   using antibody probes. The DNA sequence allows determination of the
CC   primary amino acid structure, difficult to characterise
CC   biochemically because the protein backbone is so heavily glycosyl-
CC   ated and conditions required to remove the carbohydrate tend to
CC   result in breakage of the protein.  Prods. of the DNA can be used
CC   to produce Abs useful for detecting hIM and determining glycosyl-
CC   ation patterns or glycosyltransferase activity.  They can be used
CC   for the early detection and differential diagnosis of cancers esp.
CC   epithelial cancer, and also in diseases with altered intestinal
CC   mucin prodn. such as cystic fibrosis and colitis.
CC   See also R07670-76.
SQ   Sequence   278 AA;


  Query Match             6.7%;  Score 292;  DB 1;  Length 278;
  Best Local Similarity   37.3%;  Pred. No. 2.5e-09;
  Matches   79;  Conservative   14;  Mismatches   79;  Indels   40;  Gaps    8;

Qy    502 APTAAPTAAPSTVVPPATPPATAAPVP---PTTAIPTPAPTAAPTAAPTTAAPESPTTVT 558
          |||  | :  :|| |  ||  |  |      ||   || ||   |  |||      || |
Db      6 APTPTPISTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTVL--ITTTTT 63

Qy    559 VPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAA 618
          : ||  ||:  :| |   ||| |:  || ||   |  ||  |      | || |  ||
Db     64 MTPTPTPTSTKSTTV--TPITTTTTVTATPTPTGTQTPTMIP------ISTTTTVTPTPT 115

Qy    619 PTTAAPAPNTTVTVPPTAAPTTAAPAP----------------NTTV--TVPPTAAPTA 659
          ||| :  | |  :  | |  ||: | |                :||:  |:||    |:
Db    116 PTTGSTGPPTHTSTAPIAELTTSNPPPESSTPQTSRSTSSPLTESTTLLSTLPPAIEMTS 175

Qy    660 -APPTVAHAPNTTAA-------PVTTTSAPAT 683
           |||:   || ||         | |||| | |
Db    176 TAPPSTPTAPTTTTGAHTLSPPPSTTTSPPGT 207
```

```
RESULT  13
W64359
ID   W64359 standard; Protein; 267 AA.
AC   W64359;
DT   09-NOV-1998  (first entry)
DE   Mycobacterium tuberculosis antigen TbH-29.
KW   Tuberculosis; infection; diagnosis; antigen; TbH-29.
OS   Mycobacterium tuberculosis strain H37Rv.
PN   WO9816645-A2.
PD   23-APR-1998.
PF   07-OCT-1997; U18214.
PR   13-MAR-1997; US-818111.
PR   11-OCT-1996; US-729622.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI   Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR   WPI; 98-251292/22.
DR   N-PSDB; V44403.
PT   New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT   to develop products for the detection of M. tuberculosis infection
PT   and diagnosis of tuberculosis
PS   Example 3; Page 157-158; 250pp; English.
CC   This polypeptide comprises an antigenic portion of Mycobacterium
CC   tuberculosis antigen TbH-29.  A DNA molecule (see V44403) encoding
CC   the polypeptide was isolated from a M. tuberculosis strain H37Rv
CC   genomic library.  The invention relates to compositions and methods
CC   for diagnosing tuberculosis.  It provides polypeptides (see
CC   W64291-W64379) comprising an antigenic portion of a soluble M.
CC   tuberculosis antigen, or an immunogenic portion of an M.
CC   tuberculosis antigen, as well as DNA sequences encoding such
CC   polypeptides, recombinant expression vectors and transformed or
CC   transfected host cells.  Also claimed are methods and diagnostic
CC   kits for detecting M. tuberculosis infection in a patient using
CC   these polypeptides, antibodies or oligonucleotide probes and
CC   primers, for the diagnosis of tuberculosis.
SQ   Sequence   267 AA;


  Query Match             6.6%;  Score 285.5;  DB 1;  Length 267;
  Best Local Similarity   38.8%;  Pred. No. 5.4e-09;
  Matches   81;  Conservative    5;  Mismatches  102;  Indels   21;  Gaps    8;

Qy    486 TSTAAPTAAPTA-APTAAPTAAPTAAPSTVVP-PATPPAT----AAPVPPTTAIPTPAPT 539
          |   ||   |    |||         |   ||| | | |      |:| |   |:|   |
Db     48 TQQQAPVPPPPPDDPTAGFQGGTIPAVQNVVPRPGTSPGVGGTPASPAPEAPAVPGVVPA 107

Qy    540 AAPTAAPTTA------APESPTTVTVPPTAAPTAAPTTAVPEIPIT-VTSAPTAAPTAAP 592
          |   |         |  ||  | |||   | : ||     | | ||: ||  ||
Db    108 PVPIPVPIIIPPPFPGWQPGMPTIPTAPPTTPVTTSATTPPTTPPTTPVTTPPTTPPTTPV 167

Qy    593 TAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVP 652
          |  ||  ||| |   |||| :| | |||| ||    |   | | || || |   |||
Db    168 TTPPTTPPTTPVTTPPTTV-APTTVAPTTVAP----TTVAPTTVAPATATP---TTVAPQ 219

Qy    653 PTAAPTAAPPTVAHAPNTTAAPVTTTSAP 681
          ||  ||  |         | || |   ||
Db    220 PTQQPTQQPTQQMPTQQQTVAPQTVAPAP 248


RESULT  14
W81726
ID   W81726 standard; Protein; 267 AA.
AC   W81726;
DT   27-JAN-1999  (first entry)
DE   M. tuberculosis immunogenic polypeptide TbH-29.
KW   Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW   vaccine; pharmaceutical; infection; diagnosis.
OS   Mycobacterium tuberculosis.
PN   WO9816646-A2.
PD   23-APR-1998.
PF   07-OCT-1997; U18293.
```

```
PR   13-MAR-1997; US-818112.
PR   11-OCT-1996; US-730510.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI   Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR   WPI; 98-261042/23.
DR   N-PSDB; V64512.
PT   Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT   to develop products for the detection of M. tuberculosis infection
PT   and for diagnosis, treatment and prevention of tuberculosis
PS   Example 3b; Page 149-150; 230pp; English.
CC   This sequence represents an immunogenic portion of a soluble
CC   Mycobacterium tuberculosis (MT) antigen which can be used in a method
CC   for inducing protective immunity against tuberculosis (TB). This
CC   sequence can be formulated into vaccines and/or pharmaceutical
CC   compositions for immunising against M. tuberculosis infection or may
CC   be used for the diagnosis of tuberculosis.
SQ   Sequence   267 AA;


  Query Match             6.6%;  Score 285.5;  DB 1;  Length 267;
  Best Local Similarity   38.8%;  Pred. No. 5.4e-09;
  Matches   81;  Conservative    5;  Mismatches  102;  Indels   21;  Gaps    8;

Qy     486 TSTAAPTAAPTA-APTAAPTAAPTAAPSTVVP-PATPPAT----AAPVPPTTAIPTPAPT 539
           |   ||   |    |||        |   ||| | | |      |:| |   |:|   |
Db      48 TQQQAPVPPPPPDDPTAGFQGGTIPAVQNVVPRPGTSPGVGGTPASPAPEAPAVPGVVPA 107

Qy     540 AAPTAAPTTA------APESPTTVTVPPTAAPTAAPTTAVPEIPIT-VTSAPTAAPTAAP 592
           |   |          |   ||  | |||   | : ||     | | ||: ||  ||
Db     108 PVPIPVPIIIPPPFPGWQPGMPTIPTAPPTTPVTTSATTPPTTPPTTPVTTPPTTPPTTPV 167

Qy     593 TAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVP 652
           |  ||  ||| |   |||| :| | |||| ||    |   | | || || |   |||
Db     168 TTPPTTPPTTPVTTPPTTV-APTTVAPTTVAP----TTVAPTTVAPATATP---TTVAPQ 219

Qy     653 PTAAPTAAPPTVAHAPNTTAAPVTTTSAP 681
           ||  ||  |         | || |   ||
Db     220 PTQQPTQQPTQQMPTQQQTVAPQTVAPAP 248


RESULT  15
R12535
ID   R12535 standard; Protein; 278 AA.
AC   R12535;
DT   02-SEP-1991  (first entry)
DE   SMUC-41 intestinal mucin.
KW   SMUC; intestinal mucin; diagnosis; cancer.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   region          7. .29
FT                   /label= repeat_1
FT   region          30. .52
FT                   /label= repeat_2
FT   cross_links     53. .75
FT                   /label= repeat_3
FT   region          76. .98
FT                   /label= repeat_4
FT   region          99. .121
FT                   /label= repeat_5
FT   modified_site   265. .267
FT                   /label= glycosylation_site
PN   WO9108217-A.
PD   13-JUN-1991.
PF   04-DEC-1990; U07087.
PR   05-DEC-1989; US-447140.
PA   (REGC ) UNIV OF CALIFORNIA.
PI   Kim YS, Gum JR;
DR   WPI; 91-193147/26.
DR   N-PSDB; Q12218.
PT   Nucleic acid encoding human intestinal mucin peptide(s) - used
PT   for producing prods. for diagnosis and treatment of disorders
PT   such as cancer
PS   Disclosure; Fig 2(B); 57pp; English.
CC   The sequence contains repeats having a threonine and proline-rich
CC   consensus.
CC   Antibodies produced against this mucin peptide can be used for
CC   cancer therapy by coupling to toxins and can also be used in treatment,
CC   partic. by blocking or interrupting some functional activity of
CC   intestinal mucin protein which contributes to the neoplastic phenotype.
CC   A vaccination for inducing a cellular immune response can be prepd.  The
CC   prods. are also useful as substrates for assay of glycosyltransferase
CC   activity and other glycosylation enzymes.
CC   See also Q12217-21 and Q12247-51.
SQ   Sequence   278 AA;


  Query Match             6.6%;  Score 284.5;  DB 1;  Length 278;
  Best Local Similarity   34.2%;  Pred. No. 6.3e-09;
  Matches   78;  Conservative   21;  Mismatches   96;  Indels   33;  Gaps    9;

Qy     486 TSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVP---PTTAIPTPAPTAAP 542
           |:|  ||  ||   |  ||  |    :|| |  ||  |  |      ||   || ||  |
Db      15 TTTVTPTPTPTG--TQTPTTTPITTTTTVTPTPTPTGTQTPTTVLITTTTTMTPTPT--P 70

Qy     543 TAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTT 602
           |:  :|      || ||  |  ||   |  :  |  | |  ||  ||   |  ||   |
Db      71 TSTKSTTVTPITTTTTVTATRTPTGTQTPTMIPISTTTTVTPTPTPTTGSTGPPTHTSTA 130

Qy     603 AVPEIPTTVTSPPTAAP----TTAAPAPNTTV---TVPP------TAAPTTAAPAPNTTV 649
           : |: |:   | :: |    :|::|   :|    |:||      ||||:|   || ||
Db     131 PIAELTTSNPPPESSTPQTSRSTSSPLTESTTLLSTLPPAIEMTSTAAPSTPT-APTTTS 189

Qy     650 ---TVPPTAAPTAAPPTV---------AHAPNTTAAPVTTTSAPATTP 685
              |: |  : | :||           : ||  :    |||||   ||
Db     190 GGHTLSPPPSTTTSPPGTPTRGTTTGSSSAPTPSTVQTTTTSAWTPTP 237


Search completed: July 18, 1999, 00:46:11
Job time: 264 sec
```

THIS PAGE BLANK (USPTO) COPY

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 11:36:35 ; Search time 1043.53 Seconds
(without alignments)
4123.717 Million cell updates/sec

Title: US-09-294-663-2
Perfect score: 2821
Sequence: 1 GAAAAGAATAACCAGCGAAC.....TAAAAAAAAAAAAAAAAAAA 2821

Scoring table: IDENTITY_NUC

Searched: 2002476 seqs, 762712212 residues

Database :
EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: gb_est1:*
11: gb_est2:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: gb_est6:*
16: gb_est7:*
17: gb_est8:*
18: gb_est9:*
19: gb_est10:*
20: gb_est11:*
21: gb_est12:*
22: gb_est13:*
23: gb_est22:*
24: gb_est14:*
25: gb_est15:*
26: gb_est16:*
27: gb_est17:*
28: gb_est18:*
29: gb_est19:*
30: gb_est20:*
31: gb_est21:*
32: em_est10:*
33: em_est11:*
34: em_est12:*
35: em_est13:*
36: em_est14:*
37: em_est15:*
38: em_est16:*
39: em_est17:*
40: em_est18:*
41: em_est19:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	107.6	3.8	1075	14	W29419	W29419 mb98h11.r1
c	2	101.4	3.6	538	30	AI114969	AI114969 ui41b06.y
	3	101.2	3.6	814	30	AI131654	AI131654 ui41b06.x
c	4	100.8	3.6	759	13	W28277	W28277 44g8 Human
	5	90.4	3.2	616	31	AU034820	AU034820 Dictyoste
c	6	90.2	3.2	452	22	AA589477	AA589477 vl47h12.s
c	7	90.2	3.2	514	26	AA798859	AA798859 vv94a07.r
	8	90	3.2	672	31	AU034104	AU034104 Dictyoste
c	9	89.6	3.2	482	24	AA664125	AA664125 ac06c04.s
	10	89.4	3.2	690	31	AU033402	AU033402 Dictyoste
	11	89.4	3.2	566	31	AU033450	AU033450 Dictyoste
	12	89.4	3.2	574	31	AU033594	AU033594 Dictyoste
	13	89.4	3.2	664	31	AU033617	AU033617 Dictyoste
	14	89.4	3.2	603	31	AU033659	AU033659 Dictyoste
	15	89.4	3.2	586	31	AU033708	AU033708 Dictyoste
	16	89.4	3.2	578	31	AU033712	AU033712 Dictyoste
	17	89.4	3.2	672	31	AU033776	AU033776 Dictyoste
	18	89.4	3.2	605	31	AU033778	AU033778 Dictyoste
	19	89.4	3.2	638	31	AU033787	AU033787 Dictyoste
	20	89.4	3.2	622	31	AU034132	AU034132 Dictyoste
	21	89.4	3.2	565	31	AU034176	AU034176 Dictyoste
	22	89.4	3.2	773	31	AU034671	AU034671 Dictyoste
	23	89.4	3.2	630	31	AU034708	AU034708 Dictyoste
	24	89.4	3.2	575	31	AU034754	AU034754 Dictyoste
	25	89.4	3.2	686	31	AU034976	AU034976 Dictyoste
c	26	88.8	3.1	1222	13	W42209	W42209 mc69f11.r1
	27	88.4	3.1	636	10	T55873	T55873 yb82h02.s1
c	28	88	3.1	355	26	AA798861	AA798861 vv94a10.r
	29	87.8	3.1	459	23	AI285778	AI285778 qu56b02.x
c	30	86.6	3.1	967	14	W34662	W34662 mc32c11.r1
c	31	86.2	3.1	422	25	AA673536	AA673536 vn36g11.r
	32	85.8	3.0	586	31	AU033861	AU033861 Dictyoste
	33	85.6	3.0	684	20	C25558	C25558 Dictyosteli
c	34	85	3.0	408	21	AA484400	AA484400 nf10e04.s
	35	84.4	3.0	555	31	AU033469	AU033469 Dictyoste
	36	84.4	3.0	521	31	AU033607	AU033607 Dictyoste
	37	84.2	3.0	558	31	AU033502	AU033502 Dictyoste
c	38	83.8	3.0	1461	14	W15723	W15723 mb53c01.r1
	39	83.8	3.0	576	26	AA754683	AA754683 vu20e10.r
c	40	83.6	3.0	1257	13	W98479	W98479 mg20c12.r1
c	41	83.6	3.0	1311	14	W15735	W15735 mb53g03.r1
c	42	82.8	2.9	1161	13	W41959	W41959 mc68b10.r1
	43	82.4	2.9	558	31	AU034078	AU034078 Dictyoste
	44	82	2.9	714	20	AA519116	AA519116 TgESTzz33
	45	81.8	2.9	498	31	AI253548	AI253548 aq30g07.x

ALIGNMENTS

RESULT 1
W29419/c
LOCUS W29419 1075 bp mRNA EST 11-SEP-1996
DEFINITION mb98h11.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 337509 5', mRNA sequence.
ACCESSION W29419
NID g1309566
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1075)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP

```
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: mouseest@watson.wustl.edu
           This clone is available royalty-free through LLNL ; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           MGI:218909
           Seq primer: ETPrimer
           High quality sequence stop: 64.
FEATURES             Location/Qualifiers
     source          1. .1075
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
                     /db_xref="taxon:10090"
                     /clone="337509"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>1075
BASE COUNT       37 a     82 c    789 g    159 t      8 others
ORIGIN
```

Query Match 3.8%; Score 107.6; DB 14; Length 1075;
Best Local Similarity 48.7%; Pred. No. 5.4e-12;
Matches 346; Conservative 0; Mismatches 358; Indels 6; Gaps 2;

```
Qy   1470 ACAACCAGTGCACATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTG 1529
          ||  ||    |||| |  | ||   |||| ||    | ||  |  |  |||| ||  |
Db    811 ACCCCCGCACCACAACCCCCGCCCACCCGCCACACCCCCCCGCACCCCCCCCCACATCCA 752

Qy   1530 CCCCAACGGCTGCCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCG 1589
          |||  ||  |  |||  || || ||  ||  ||||      |||||  |  | ||| ||
Db    751 CCCACACCCCACCCCTCACCGCAGCCACCCTCACCCCTCGTCCACCCTCCCCCCCAACCC 692

Qy   1590 CAACTGCAGCCCCAGTC-----CCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGC 1644
          | || ||||| ||    |     || |||   ||| ||    || | |    | || ||
Db    691 CCACCGCAGCTCCCCGCGCGAGCCCCCTCACACCCCAGCCACTCCACGACACACCCCCCA 632

Qy   1645 TGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACC 1704
            | | |    |  | ||     |  | |||||     || ||  | |    |||   ||
Db    631 CCCACACCAACCCCCCCCAGACCCCACCGCCCCCCGCGCCACACACCCCCCAACACCCCC 572

Qy   1705 ACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGT 1764
            ||  || |  | |||  |||  ||||  |  | |  |    ||      ||| | | |
Db    571 CCCGCCTCCCCCGCCCCACGCCACCCCACCGCCCCCGCCACCCTCCCCCCCCCACCCCCCC 512

Qy   1765 CACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCC 1824
          | |  |  | ||  || |  | | | ||     |||| || |  ||| |  |||  ||||
Db    511 CCCGCCCCCACCACCCCCGCCACACCCCCGCAACCCCCCCACCCCCCACCACGCCACCCC 452

Qy   1825 TACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCAC 1884
            |  |       ||       ||      |   | |  | |  ||  |  | || ||| |
Db    451 ACCCGCACACCCACCNCNNNCGCCCGGCCCACGCCCCCCCCCACCCCGCACCGGCACCCGC 392

Qy   1885 TACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCTACTAC 1944
           ||| |  | ||  | | |  | |   || |   | ||||  |  ||||  |  |
Db    391 CACCCCCCCCCCCCCCACACCTCCACCCCACCAACGACACCCCCAC-CCGCCCCCCCCCCC 333

Qy   1945 CGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTACCGC 2004
          |||| | || || |||| | || |     |  | || ||   |||  ||||  |  |  |
Db    332 CGCACCCCCCGCCCCCACCCCCCCCACCCCCCCCCCCCCCGCGCGCCCCCCCCCCCCCCC 273
```

```
Qy   2005 AGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCC 2064
            | ||   ||||| | || |   | |    |||||||  |  ||||| | ||  | ||
Db    272 CCCCCCCCACCCCACCCCCCCCCCCCCCCCCCCCCACCCACCCCCCCCCCCCCCGCCCCCC 213

Qy   2065 CCCTACCGTCGCACCTGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACC 2124
          |||  ||  | | || || ||   | || |  |  ||||        |  ||  | | ||
Db    212 CCCCCCCCCCCCCCCCCGCCCCACCCCCCCCCACCCCCCCACCCCACCCCCCCCCCCCCC 153

Qy   2125 AGCTACCACACCTGAAGATGATGACATCGACCCCCCTCTCCCCAACGACC 2174
            |  || | || |      |    |  | |||| || |  |||  |  ||
Db    152 CCCCCCCCCCCCCGCCCCCGGCCCCCCCCACCCACCCCCACCCCCCCCC 103
```

```
RESULT   2
AI114969/c
LOCUS       AI114969      538 bp    mRNA            EST       02-SEP-1998
DEFINITION  ui41b06.y1 Sugano mouse embryo mewa Mus musculus cDNA clone 1884947
            5' similar to TR:O14633 O14633 SKIN-SPECIFIC PROTEIN. ;, mRNA
            sequence.
ACCESSION   AI114969
NID         g3515293
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 538)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:969271
            Seq primer: custom primer used
            High quality sequence stop: 490.
FEATURES             Location/Qualifiers
     source          1. .538
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a  DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert.  Size selection was performed to exclude fragments
                     <1.5kb.  Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science).
                     Custom primers  for sequencing: 5' end primer
                     CTTCTGCTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTCGAGCACA."
                     /db_xref="taxon:10090"
                     /clone="1884947"
                     /clone_lib="Sugano mouse embryo mewa"
                     /dev_stage="embryo, 14 dpc"
                     /lab_host="DH10B"
BASE COUNT       91 a    149 c    190 g    108 t
ORIGIN
```

```
  Query Match             3.6%;  Score 101.4;  DB 30;  Length 538;
  Best Local Similarity   55.1%;  Pred. No. 9.1e-11;
  Matches  220;  Conservative    0;  Mismatches  176;  Indels    3;  Gaps    1;

Qy   1629 CTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAA 1688
          ||   ||  |||| |||||  ||||| | ||| |  | |||||| || | | | |  |
Db    525 CTGTTGCTGCCACTGCTGCTGCCACCACTGCTGCCCCCACTGCTACCACAGCAGCTGCTG 466

Qy   1689 CCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTG 1748
          |||| |  || |   |  |||||||  |  |  | || ||| |  | |||||||| | | |
Db    465 CCACCGCTACTGCCACCGCTACTGCCACCACTGCTGCTGCCACAGCAACTGCTGCCACCG 406

Qy   1749 AAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTG 1808
               |  ||| |  ||  | ||  |  | |||||| |  |  ||||  |||||||||
Db    405 CTGCTGCTGCCACAGCAACTGCTGCCACCGCTGCTGCCACAGCAACTGCTGCCACCGCTG 346

Qy   1809 CCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCAC 1868
          |  ||||||| ||  | ||  |||||| | ||| ||   |  | || |   |  |  ||
Db    345 CTGCCACCGCTGCTGCCACCGCTGCTGCCACAGCAACTGCTGCCACCGCTGCTGCCACAG 286

Qy   1869 CTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCA 1928
          | |||||||| || || |  ||||| |     |    ||  | |  || |    |
Db    285 CAACTGCTGCCACCGCTGCTACAGCATCCAGAGCTGTGACGTCGGCGACGGAGAGA---G 229

Qy   1929 CTGCTGCCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCG 1988
            | |  | |       ||     | ||  ||| |||  |     ||||| || || |||
Db    228 AATCGGGGCTTGTGGTGGCTCAGACAGCAGCCACCACATCCAGAGCTGCAGCAGCCCCCG 169

Qy   1989 CAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAG 2027
           |||    || || ||||| ||     || |  | ||||
Db    168 GAGCTGGAACCACAGCAGCCCCCAGAACCCAGGCTACAG 130


RESULT   3
AI131654
LOCUS       AI131654       814 bp    mRNA            EST       14-SEP-1998
DEFINITION  ui41b06.x1 Sugano mouse embryo mewa Mus musculus cDNA clone 1884947
            3' similar to TR:O14633 O14633 SKIN-SPECIFIC PROTEIN. ;, mRNA
            sequence.
ACCESSION   AI131654
NID         g3601670
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 814)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:969271
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 375.
FEATURES             Location/Qualifiers
     source          1. .814
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a  DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert.  Size selection was performed to exclude fragments
                     <1.5kb.  Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science).
                     Custom primers  for sequencing: 5' end primer
                     CTTCTGCTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTCGAGCACA."
                     /db_xref="taxon:10090"
                     /clone="1884947"
                     /clone_lib="Sugano mouse embryo mewa"
                     /dev_stage="embryo, 14 dpc"
                     /lab_host="DH10B"
BASE COUNT      184 a    245 c    211 g    173 t      1 others
ORIGIN


  Query Match             3.6%;  Score 101.2;  DB 30;  Length 814;
  Best Local Similarity   55.0%;  Pred. No. 9.8e-11;
  Matches  221;  Conservative    0;  Mismatches  178;  Indels    3;  Gaps    1;

Qy   1629 CTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAA 1688
          ||   ||  |||| |||||  ||||| | ||| |  | |||||| || | | | |  |
Db    337 CTGTTGCTGCCACTGCTGCTGCCACCACTGCTGCCCCCACTGCTACCACAGCAGCTGCTG 396

Qy   1689 CCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTG 1748
          |||| |  || |   |  ||||||  |  |  | || ||| |  | |||||||| | | |
Db    397 CCACCGCTACTGCCACCGTTACTGCCACCACTGCTGCTGCCACAGCAACTGCTGCCACCG 456

Qy   1749 AAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTG 1808
               |||  |  ||  | ||  |  | |||||| |  |  ||||  |||||||||
Db    457 CTGCTGCTGCCACAGCAACTGCTGCCACCGCTGCTGCCACAGCAACTGCTGCCACCGCTG 516

Qy   1809 CCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCAC 1868
          |  ||||||| ||  | ||  |||||| | ||| ||   |  | || |   |  |  ||
Db    517 CTGCCACCGCTGCTGCCACCGCTGCTGCCACAGCAACTGCTGCCACCGCTGCTGCCACAG 576

Qy   1869 CTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCA 1928
          | |||||||| || || |  ||||| |     |    ||  | |  || |    |
Db    577 CAACTGCTGCCACCGCTGCTACAGCATCCAGAGCTGTGACGTCGGCGACGGAGAGA---G 633

Qy   1929 CTGCTGCCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCG 1988
            | |  | |       ||     | ||  ||| |||  |     ||||| || || |||
Db    634 AATCGGGGCTTGTGGTGGCTCAGACAGCAGCCACCACATCCAGAGCTGCAGCAGCCCCCG 693

Qy   1989 CAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCA 2030
           |||    || || ||||| ||     || |  | ||||  ||
Db    694 GAGCTGGAACCACAGCAGCCCCCAGAACCCAGGCTACAGCCA 735


RESULT   4
W28277/c
LOCUS       W28277         759 bp    mRNA            EST       08-MAY-1996
DEFINITION  44g8 Human retina cDNA randomly primed sublibrary Homo sapiens
            cDNA, mRNA sequence.
ACCESSION   W28277
NID         g1308243
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 759)
  AUTHORS   Macke, J., Smallwood, P. and Nathans, J.
  TITLE     Adult Human Retina cDNA
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Dr. Jeremy Nathans
            Dr. Jeremy Nathans, Dept. of Molecular Biology and Genetics
```

```
            Johns Hopkins School of Medicine
            725 North Wolfe Street, Baltimore, MD 21205
            Tel: 410 955 4678
            Fax: 410 614 0827
            Email: jeremy_nathans@qmail.bs.jhu.edu
            Clones from this library are NOT available.
            PCR PRimers
            FORWARD: CTTTTGAGCAAGTTCAGCCTGGTTAAGT
            BACKWARD: GAGGTGGCTTATGAGTATTTCTTCCAGGGTAA
            Seq primer: GGGTAAAAAGCAAAAGAATT.
FEATURES              Location/Qualifiers
     source           1. .759
                      /organism="Homo sapiens"
                      /note="Organ: eye; Vector: lambda gt10; Site_1: EcoRI;
                      Site_2: EcoRI; The library used for sequencing was a
                      sublibrary derived from a human retina cDNA library.
                      Inserts from retina cDNA library DNA were isolated,
                      randomly primed, PCR amplified, size-selected, and cloned
                      into lambda gt10.  Individual plaques were arrayed and
                      used as templates for PCR amplification, and these PCR
                      products were used for sequencing."
                      /db_xref="taxon:9606"
                      /clone_lib="Human retina cDNA randomly primed sublibrary"
                      /sex="mixed (males and females)"
                      /tissue_type="retina"
                      /dev_stage="adult"
                      /lab_host="E. coli strain K802"
     mRNA             <1. .>759
BASE COUNT       38 a     71 c    494 g     16 t    140 others
ORIGIN


  Query Match             3.6%;  Score 100.8;  DB 13;  Length 759;
  Best Local Similarity   41.8%;  Pred. No. 1.2e-10;
  Matches  272;  Conservative    0;  Mismatches  377;  Indels    1;  Gaps    1;

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
           ||||  |  ||   ||| || |||  |  || |  ||  |||| ||| |  | ||  ||
Db    732 CCCCCCCGNCCNNCCCCCCCCCCGCNGCGCCNCGCGCCCCGCCNCCCCCNNCCGCCNCCG 673

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
           |   |   ||     || |           ||  ||| | |||| ||||||  |||| || |
Db    672 CNCCCTNCCCCCGCNCCCCNCCCGNCCGCCCCCCCCNCCCCGCGGCCCCCCCCGCNGCCC 613

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
             |  |  |   |||     ||||   | |   |          |  |  |  | | |
Db    612 GCNCGCCNCCGCGCCCCCCCNCCCCNNNCCCCCCCNGCCCNANCNCCCCCCCCNCNCNCC 553

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
            | |   |||  |    ||   | |     |||   | | | | |  |  | ||  | |
Db    552 GCCCCCGCCCNGCCCNCGC-NCCGCNCCCNCCCCNCNCGCGGCCCCCCCNCCCCCCCCNG 494

Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
             ||||  || |    ||| ||    ||||  || | | |||     | || | ||
Db    493 GCCCCCCNCCCCCGCNNCCCCCCCCGCNCCCCNCCCCCCGNCCCCNGGGCGGCNGCCCNCC 434

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
            ||  |  |   |    | || ||  |  |  |||| | ||  |  | |  ||   ||
Db    433 CNCCNCCCCCCCCCCCCNNCCCCCCCCCCGCCCCCCCCCCNCCCACNCCCCCCCCCNGCC 374

Qy   1902 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCA 1961
           ||  | || |    |    || ||  |    ||||  |   | |  | ||  | |||
Db    373 CCCNCCCCCCCCNNNCCNCCGCCCCCNCCCNNNCCCCCCNCGCGCCCCCCCCCCCCCCCCC 314

Qy   1962 ACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACA 2021
            | |         |  | || ||| || |       |  |           |       | |
Db    313 CCCCNCNCCCCCCCCCCCCCNCCCNCCNCNNTTTTTTTTTTTTTNTTTCNNGGNGNGGNAAA 254

Qy   2022 CCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTG 2081
            || | |     |    |  | || || | ||  | |||||  || || | |||
Db    253 ANNNAGNCTCCCNCNCCGNNNCGCCCCCCNCCCCCGCCNCCCCCCCCCCCCTCNCCCCTC 194

Qy   2082 CACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAG 2141
```

```
          | |||  | |  |     ||||     |  |||    | | ||  |  ||    |
Db    193 CCCCCCCCNCGCCCCNCCCCCCNCCCCCCCCCAAAACCCCCCCCCCCCCCCCCGCNCCCCCC 134

Qy   2142 ATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTT 2191
                |  |  | |||| | ||||  |              ||   ||
Db    133 CCNNNCCCCNCCCCNCCCCCCCCCCCCCCCCNNNNNCNNNNNNNTTTTTTT 84


RESULT   5
AU034820
LOCUS       AU034820      616 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLE543, mRNA sequence.
ACCESSION   AU034820
NID         g3800244
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 616)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES              Location/Qualifiers
     source           1. .616
                      /organism="Dictyostelium discoideum"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone_lib="SL"
                      /dev_stage="slug"
BASE COUNT      185 a    194 c     55 g    181 t      1 others
ORIGIN


  Query Match             3.2%;  Score 90.4;  DB 31;  Length 616;
  Best Local Similarity   50.3%;  Pred. No. 1.3e-08;
  Matches  220;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;

Qy   1482 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1541
           | ||  |||||  |  | || |  | |||||      |  || |  |  |  |||
Db     44 CTTCTGCTGCTTCATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTG 103

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
           |  |||    |  | || ||  |    ||    |  | ||| |  || |||||  |
Db    104 CTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCAT 163

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
           ||  | |  |  || |  ||  | |  |  |  | |  || |  |  |  | ||| |
Db    164 CAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCAT 223

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
           | |   |||||  ||  |||   ||    | |  ||  |  | | || || |  ||||
Db    224 CAAGCTCTGCTCCATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCAT 283

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
           | |||||  ||||  ||||     |         | |   |||  ||  || |  | ||
Db    284 CTGCAGCAACTACAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTA 343

Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
           |  || | ||| | ||  | ||| | ||| | ||| ||||  | || || ||||   |
Db    344 CAACCGCAACCACAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCA 403

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
           |    |||| || |   || |  || || |  || |  || | ||| ||||  |  ||
```

```
Db    404 TAGNTACAACCACAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCA 463

Qy   1902 CCAACACCACAGTCACT 1918
          |    || ||    |||
Db    464 CAGCTACAACCACAACT 480


RESULT   6
AA589477/c
LOCUS       AA589477      452 bp    mRNA            EST       16-SEP-1997
DEFINITION  vl47h12.s1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            975431 3', mRNA sequence.
ACCESSION   AA589477
NID         g2402857
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:556159
            Possible reversed clone: polyT not found
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:10090"
                     /clone="975431"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
     mRNA            complement(<1. .>452)
BASE COUNT       32 a     93 c    193 g    134 t
ORIGIN


  Query Match             3.2%;  Score 90.2;  DB 22;  Length 452;
  Best Local Similarity   54.3%;  Pred. No. 1.4e-08;
  Matches  182;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy   1503 CTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTCCA 1562
          ||  | |  |  | ||||   |  | ||| || | || ||| ||   ||| || || |
Db    346 CTTTTGCTGCAACAGCCCTTCCCACAGCCACAGCCGCAGCCACAGGAGCCACAGCCACAG 287

Qy   1563 CCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAA 1622
             | ||   || | ||  || || | |||       |    |  || |  | || |||

Db    286 GAGCGGCAGCAGCAGCGGCGGCAGCAGCAGACTACCACTGGCTTGCAGCAGCCACAGCAG 227

Qy   1623 TTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAAT 1682
             |  |  |  || | || ||  |  ||||| ||||  | || ||  | ||| | | |
Db    226 CCACAGCAGCCACCACAACAGCCACAGCCACCACAGCCACCACAACCACAGCCACCGCAG 167

Qy   1683 CCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCCTACTACTGCTG 1742
          || || |||| |    |   | |  | || ||  |  ||||| ||||| |  | || || |
Db    166 CCACAGCCACCGCAGCCACAGCCACCACAGCCACAGCCACCACAGCCACAGCCACCGCAG 107

Qy   1743 TCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCA 1802
          | | |  | | ||  |||   |||  ||||  |  | ||||||  |  ||||  |  |||
Db    106 CCACAGCCACCACAGCCACAACCACCACAGCCACCGCAGCTGCCACAGCCGCCACAGCCA 47

Qy   1803 CCGCTGCCCCCACCGCAGCCCCCTACTACTGCTGTC 1837
          || |  || |||| ||  |  | ||   || ||||
Db     46 CCACAACCTCCACAGCCACAGCAACCCATGTTGTC 12


RESULT   7
AA798859/c
LOCUS       AA798859      514 bp    mRNA            EST       10-FEB-1998
DEFINITION  vv94a07.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            1230036 5', mRNA sequence.
ACCESSION   AA798859
NID         g2861814
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:655628
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .514
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:10090"
                     /clone="1230036"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
BASE COUNT       41 a    111 c    207 g    155 t
ORIGIN
```

```
  Query Match             3.2%;  Score 90.2;  DB 26;  Length 514;
  Best Local Similarity   54.3%;  Pred. No. 1.4e-08;
  Matches  182;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy   1503 CTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTCCA 1562
          ||  | |  |  | ||||   |  | ||| || | || ||| ||   ||| || || |
Db    352 CTTTTGCTGCAACAGCCCTTCCCACAGCCACAGCCGCAGCCACAGGAGCCACAGCCACAG 293

Qy   1563 CCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAA 1622
            | ||   || | ||  || || | |||        |    |  || |  | || |||
Db    292 GAGCGGCAGCAGCAGCGGCGGCAGCAGCAGACTACCACTGGCTTGCAGCAGCCACAGCAG 233

Qy   1623 TTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAAT 1682
            |  |  |  || | || ||  |  ||||| |||| | || ||  | ||| | | |
Db    232 CCACAGCAGCCACCACAACAGCCACAGCCACCACAGCCACCACAACCACAGCCACCGCAG 173

Qy   1683 CCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTG 1742
          || || |||| |   |   | |  | || ||  |  ||||| ||||| |  | || || |
Db    172 CCACAGCCACCGCAGCCACAGCCACCACAGCCACAGCCACCACAGCCACAGCCACCGCAG 113

Qy   1743 TCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCA 1802
            | |  | | ||  |||   |||  |||| |  | ||||| |  ||||  |  |||
Db    112 CCACAGCCACCACAGCCACAACCACCACAGCCACCGCAGCTGCCACAGCCGCCACAGCCA 53

Qy   1803 CCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTC 1837
          || |  || |||| ||  |  | ||   || ||||
Db     52 CCACAACCTCCACAGCCACAGCAACCCATGTTGTC 18


RESULT   8
AU034104
LOCUS       AU034104      672 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLB746, mRNA sequence.
ACCESSION   AU034104
NID         g3799528
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 672)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 672)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      214 a     207 c      60 g     190 t       1 others
ORIGIN


  Query Match             3.2%;  Score 90;  DB 31;  Length 672;
  Best Local Similarity   50.1%;  Pred. No. 1.6e-08;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1482 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1541
          ||||  |  |  |  |  | |||||     |  || |  |  |  ||||  |||   ||
Db     78 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 137

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
          |  ||| || |  |  |  | | |     |||| || || |  || |  |||   |  |||
Db    138 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 197

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
          ||    || || || |  |||  || | ||  |  |  | ||  |  | | | | ||||
Db    198 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 257

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
          |  || |  ||||  |    ||  |  |  |      ||    || |  |||| ||  | |
Db    258 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 317

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
          | || ||  |||| | |||      | |       | |  ||||  ||  || |  | |||
Db    318 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 377

Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
          | ||  | ||| | ||| | ||| | ||  | ||| ||||  | || |  |||    |
Db    378 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 437

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
            |    |||||||     || |  || |  |  |  |  | || ||| ||||  |  ||
Db    438 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 497

Qy   1902 CCAACACCACAGTCACTGTACCA 1924
          |    ||| |    | |  | ||
Db    498 CANCTACCTCTTAAATTTAATCA 520


RESULT   9
AA664125/c
LOCUS       AA664125      482 bp    mRNA            EST       12-NOV-1997
DEFINITION  ac06c04.s1 Stratagene lung (#937210) Homo sapiens cDNA clone 855654
            3' similar to contains element MSR1 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AA664125
NID         g2618116
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Homo sapiens"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. normal lung.  Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:9606"
                     /clone="855654"
                     /clone_lib="Stratagene lung (#937210)"
                     /sex="male"
                     /dev_stage="72 years"
                     /lab_host="SOLR cells (kanamycin resistant)"
```

```
BASE COUNT       51 a    113 c    176 g    142 t
ORIGIN

  Query Match             3.2%;  Score 89.6;  DB 24;  Length 482;
  Best Local Similarity   50.0%;  Pred. No. 1.9e-08;
  Matches  224;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy   1485 CGACTGCTGCCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCC 1544
          | || || ||  |  |||| ||  |||| || ||     | |   |  ||  |
Db    459 CTACAGCAGCAGCAGCAGCAGCAACAGCAGCAGTTCCAGGCTCAGCAGAGTGCCATG 400

Qy   1545 CAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAG 1604
          || | || |  ||   ||     |  || | ||| |  | || ||| | |||||  |||
Db    399 CAGCAGCAGTTCCAAGCAGTAGTGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCAG 340

Qy   1605 TCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTA 1664
            || |  | |    |||| |  |  |     |  |  | |   |  | |||||  |
Db    339 CAGCAGCATCTAATTAAATTGCATCATCAAAATCAGCAACAGATACAGCAGCAGCAACAG 280

Qy   1665 CTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCG 1724
          |  |||| ||   |    |  | |  ||   ||| | || |  | || ||  |  | |
Db    279 CAGCTGCAGCGAATAGCACAGCTGCAGCTCCAACAACAGCAACAGCAGCAGCAGCAGCAG 220

Qy   1725 CAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTG 1784
          ||||  |  |  | || | |  || |  ||||  |||     || ||||  |    || |
Db    219 CAGCAGCAGCAGCAGGCTTTGCAGGCCCAGCCACCAATTCAGCAGCCACCGATGCAG 160

Qy   1785 CCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAA 1844
          |  |||| ||  || ||  | | | | |  || ||||  || |  | | ||   || | |
Db    159 CAGCCACAGCCTCCGCCCTCCCAGGCTCTGCCCCAGCAGCTGCAGCAGATGCATCACACA 100

Qy   1845 TCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCA 1904
            || |  | | | |  |  || |  | || ||| ||| |  | ||| |||  ||  ||
Db     99 CAGCACCACCAGCCGCCACCACAGCCCCAGCAGCCTCCAGTTGCTCAGAACCAACCATCA 40

Qy   1905 ACACCACAGTCACTGTACCACCCACTGC 1932
            ||  | | ||| ||  ||  | | ||
Db     39 CAACTCCCGCCACAGTCGCAGACCCAGC 12


RESULT  10
AU033402
LOCUS       AU033402      690 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLA726, mRNA sequence.
ACCESSION   AU033402
NID         g3798826
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 690)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      212 a    213 c     67 g    198 t
ORIGIN

  Query Match             3.2%;  Score 89.4;  DB 31;  Length 690;
  Best Local Similarity   50.1%;  Pred. No. 2.1e-08;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1482 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1541
          ||||  |  |  |  |  | |||||     |  || |  |  |  ||||  |||   ||
Db    100 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 159

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
          | ||| || |  |  |  | | |     |||| || || |  || |  |||   |  |||
Db    160 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 219

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
          ||    || || || |  |||   || | ||  |  |  | ||  |  | | | | ||||
Db    220 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 279

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
          |  || |  |||| |    ||  |  | |     ||    || |  |||| ||  | |
Db    280 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 339

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
          | || ||  |||| | |||     | |       | |  ||||  ||  || | | |||
Db    340 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 399

Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
          | ||  | ||| | ||| | ||| | ||  | ||| ||||  | || |  |||    |
Db    400 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 459

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
           |    |||||||     || |  || | |  |  |  | || ||| ||||  |  ||
Db    460 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 519

Qy   1902 CCAACACCACAGTCACTGTACCA 1924
          |    ||| |    | |  | ||
Db    520 CAGCTACCTCTTAAATTTAATCA 542


RESULT  11
AU033450
LOCUS       AU033450      566 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLA841, mRNA sequence.
ACCESSION   AU033450
NID         g3798874
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 566)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 566)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .566
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      182 a    179 c     46 g    159 t
ORIGIN
```

```
  Query Match             3.2%;  Score 89.4;  DB 31;  Length 566;
  Best Local Similarity   50.1%;  Pred. No. 2.1e-08;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1482 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1541
          ||||  |  |  |  |  | |||||     |  || |  |  |  ||||  |||   ||
Db      6 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 65

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
          | ||| || |  |  |  | | |     |||| || || |  || |  |||   |  |||
Db     66 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 125

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
          ||    || || || |  |||   || | ||  |  |  | ||  |  | | | | ||||
Db    126 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 185

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
          |  || |  |||| |    ||  |  | |     ||    || |  |||| ||  | |
Db    186 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 245

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
          | || ||  |||| | |||     | |      | |  ||||  ||  || |  | |||
Db    246 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 305

Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
          | ||  | ||| | ||| | ||| | ||  | ||| ||||  | || |  |||    |
Db    306 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 365

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
          |    |||||||     || |  || |  |  |  |  | || ||| ||||  |  ||
Db    366 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 425

Qy   1902 CCAACACCACAGTCACTGTACCA 1924
          |    ||| |    | |  | ||
Db    426 CAGCTACCTCTTAAATTTAATCA 448


RESULT  12
AU033594
LOCUS       AU033594      574 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLB210, mRNA sequence.
ACCESSION   AU033594
NID         g3799018
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 574)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      182 a    182 c     47 g    162 t      1 others
ORIGIN
```

```
  Query Match             3.2%;  Score 89.4;  DB 31;  Length 574;
  Best Local Similarity   50.1%;  Pred. No. 2.1e-08;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1482 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1541
          ||||  |  |  |  |  | |||||     |  || |  |  |  ||||  |||   ||
Db     14 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 73

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
          | ||| || |  |  |  | | |     |||| || || |  || |  |||   |  |||
Db     74 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 133

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
          ||    || || || |  |||   || | ||  |  |  | ||  |  | | | | ||||
Db    134 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 193

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
          |  || |  |||| |    ||  |  | |     ||    || |  |||| ||  | |
Db    194 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 253

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
          | || ||  |||| | |||     | |      | |  ||||  ||  || |  | |||
Db    254 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 313

Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
          | ||  | ||| | ||| | ||| | ||  | ||| ||||  | || |  |||    |
Db    314 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 373

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
          |    |||||||     || |  || |  |  |  |  | || ||| ||||  |  ||
Db    374 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 433

Qy   1902 CCAACACCACAGTCACTGTACCA 1924
          |    ||| |    | |  | ||
Db    434 CAGCTACCTCTTAAATTTAATCA 456


RESULT  13
AU033617
LOCUS       AU033617      664 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLB234, mRNA sequence.
ACCESSION   AU033617
NID         g3799041
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 664)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax;+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .664
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      193 a    213 c     66 g    192 t
ORIGIN
```

```
  Query Match             3.2%;  Score 89.4;  DB 31;  Length 664;
  Best Local Similarity   50.1%;  Pred. No. 2.1e-08;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1482 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1541
          ||||  |  |  |  |  | |||||     |  || |  |  |  ||||  |||   ||
Db    103 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 162

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
          | ||| || |  |  |  | | |     |||| || || |  || |  |||   |  |||
Db    163 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 222

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
          ||    || || || |  |||   || | ||  |  |  | ||  |  | | | | ||||
Db    223 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 282

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
          |  || |  |||| |    ||  |  |  |     ||    || |  |||| ||  | |
Db    283 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 342

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
          | || ||  |||| | |||     | |       | |  ||||  ||  || |  | |||
Db    343 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 402

Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
          | ||  | ||| | ||| | ||| | ||  | ||| ||||  | || |  |||    |
Db    403 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 462

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
            |    |||||||     || |  || |  |  | |  |  | || ||| ||||  |  ||
Db    463 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 522

Qy   1902 CCAACACCACAGTCACTGTACCA 1924
          |    ||| |    | |  | ||
Db    523 CAGCTACCTCTTAAATTTAATCA 545


RESULT  14
AU033659
LOCUS       AU033659      603 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLB278, mRNA sequence.
ACCESSION   AU033659
NID         g3799083
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 603)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      184 a    192 c     53 g    174 t
ORIGIN


  Query Match             3.2%;  Score 89.4;  DB 31;  Length 603;
  Best Local Similarity   50.1%;  Pred. No. 2.1e-08;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1482 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1541
          ||||  |  |  |  |  | |||||     |  || |  |  |  ||||  |||   ||
Db     43 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 102

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
          | ||| || |  |  |  | | |     |||| || || |  || |  |||   |  |||
Db    103 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 162

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
          ||    || || || |  |||   || | ||  |  |  | ||  |  | | | | ||||
Db    163 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 222

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
          |  || |  |||| |    ||  |  |  |     ||    || |  |||| ||  | |
Db    223 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 282

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
          | || ||  |||| | |||     | |       | |  ||||  ||  || |  | |||
Db    283 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 342

Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
          | ||  | ||| | ||| | ||| | ||  | ||| ||||  | || |  |||    |
Db    343 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 402

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
            |    |||||||     || |  || |  |  | |  |  | || ||| ||||  |  ||
Db    403 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 462

Qy   1902 CCAACACCACAGTCACTGTACCA 1924
          |    ||| |    | |  | ||
Db    463 CAGCTACCTCTTAAATTTAATCA 485


RESULT  15
AU033708
LOCUS       AU033708      586 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLB343, mRNA sequence.
ACCESSION   AU033708
NID         g3799132
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 586)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      182 a    186 c     50 g    168 t
ORIGIN


  Query Match             3.2%;  Score 89.4;  DB 31;  Length 586;
  Best Local Similarity   50.1%;  Pred. No. 2.1e-08;
```

Matches 222; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

```
Qy   1482 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1541
          ||||  |  |  |  |  | |||||      |  || |  |  |  ||||  |||   ||
Db     26 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 85

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
          | ||| || |  |  |  | | |     |||| || || |  || |  |||   |  |||
Db     86 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 145

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
          ||    || || || |  |||   || | ||  |  |  | ||  |  | | | | ||||
Db    146 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 205

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
          |  || |  |||| |    ||  |  | |     ||    || |  |||| ||  | |
Db    206 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 265

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
          | || ||  |||| | |||     | |      | |  ||||  ||  || |  | |||
Db    266 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 325

Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
          | ||  | ||| | ||| | ||| | ||  | ||| ||||  | || |  |||    |
Db    326 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 385

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
            |    |||||||     || |  || |  |  |  |  | || ||| ||||  |  ||
Db    386 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 445

Qy   1902 CCAACACCACAGTCACTGTACCA 1924
          |    ||| |    | |  | ||
Db    446 CAGCTACCTCTTAAATTTAATCA 468
```

Search completed: July 18, 1999, 11:36:45
Job time: 1753 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 02:08:42 ; Search time 1514.49 Seconds
(without alignments)
4140.380 Million cell updates/sec

Title: US-09-294-663-2
Perfect score: 2821
Sequence: 1 GAAAAGAATAACCAGCGAAC.....TAAAAAAAAAAAAAAAAAAA 2821

Scoring table: IDENTITY_NUC

Searched: 3232953 seqs, 1111407957 residues

Database : Pending_Patents_NA:*
1: /cgn2_6/ptodata/1/pna/PCTUS9_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US081A_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US081B_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US081C_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US081D_COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US082A_COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US082B_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US082C_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US083A_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US083B_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US084A_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US084B_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US084C_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US087A_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US087B_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US087C_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US088A_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US088B_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US088C_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US088D_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US089A_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US089B_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US089C_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US089D_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US089E_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US090A_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US090B_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US090C_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US091A_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US091B_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US092A_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US092B_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US092C_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US092D_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US6001_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US6002A_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US6002B_COMB.seq:*
44: /cgn2_6/ptodata/1/pna/US6003A_COMB.seq:*
45: /cgn2_6/ptodata/1/pna/US6003B_COMB.seq:*
46: /cgn2_6/ptodata/1/pna/US6004A_COMB.seq:*
47: /cgn2_6/ptodata/1/pna/US6004B_COMB.seq:*
48: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*
49: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*
50: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*
51: /cgn2_6/ptodata/1/pna/US6008A_COMB.seq:*
52: /cgn2_6/ptodata/1/pna/US6008B_COMB.seq:*
53: /cgn2_6/ptodata/1/pna/US6008C_COMB.seq:*
54: /cgn2_6/ptodata/1/pna/US6009A_COMB.seq:*
55: /cgn2_6/ptodata/1/pna/US6009B_COMB.seq:*
56: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq:*
57: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq:*
58: /cgn2_6/ptodata/1/pna/US6011_COMB.seq:*
59: /cgn2_6/ptodata/1/pna/US6012_COMB.seq:*
60: /cgn2_6/ptodata/1/pna/US6013_COMB.seq:*
61: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
62: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
63: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
64: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
65: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2821	100.0	2821	33	US-09-103-429-2	Sequence 2, Appli
	2	2821	100.0	2821	33	US-09-103-429A-2	Sequence 2, Appli
	3	2821	100.0	2821	38	US-09-294-663-2	Sequence 2, Appli
	4	2354.4	83.5	2455	33	US-09-103-429-1	Sequence 1, Appli
	5	2354.4	83.5	2455	33	US-09-103-429A-1	Sequence 1, Appli
	6	2354.4	83.5	2455	38	US-09-294-663-1	Sequence 1, Appli
	7	158.8	5.6	8705	1	PCT-US99-06742-14	Sequence 14, Appl
c	8	158.8	5.6	10596	13	US-08-407-128-15	Sequence 15, Appl
c	9	158.8	5.6	10596	13	US-08-409-770-15	Sequence 15, Appl
c	10	158.8	5.6	10596	15	US-08-487-827-15	Sequence 15, Appl
c	11	158.8	5.6	10596	15	US-08-487-831-15	Sequence 15, Appl
c	12	158.8	5.6	803	18	US-08-733-369A-104	Sequence 104, App
c	13	158.8	5.6	799	18	US-08-733-369A-105	Sequence 105, App
c	14	158.8	5.6	798	18	US-08-733-369A-112	Sequence 112, App
c	15	158.8	5.6	9600	25	US-08-910-647-1	Sequence 1, Appli
c	16	158.8	5.6	798	28	US-08-970-900-4	Sequence 4, Appli
c	17	158.8	5.6	799	28	US-08-970-900-6	Sequence 6, Appli
c	18	158.8	5.6	803	28	US-08-970-900-7	Sequence 7, Appli
c	19	158.8	5.6	2580	31	US-09-050-863-2	Sequence 2, Appli
	20	158.8	5.6	5452	33	US-09-130-114-1	Sequence 1, Appli
c	21	158.8	5.6	10380	33	US-09-130-115-1	Sequence 1, Appli
c	22	147	5.2	947	59	US-60-129-391-160	Sequence 160, App
	23	137.8	4.9	5163	1	PCT-US97-14104-1	Sequence 1, Appli
	24	137.8	4.9	5318	1	PCT-US97-14104-2	Sequence 2, Appli
	25	137.8	4.9	5163	18	US-08-700-651A-1	Sequence 1, Appli
	26	137.8	4.9	5318	18	US-08-700-651A-2	Sequence 2, Appli
	27	137.8	4.9	5163	18	US-08-700-651-1	Sequence 1, Appli
	28	137.8	4.9	5318	18	US-08-700-651-2	Sequence 2, Appli
	29	137.8	4.9	5318	26	US-08-928-361-3	Sequence 3, Appli
	30	137.8	4.9	5163	26	US-08-928-361-4	Sequence 4, Appli
	31	137.8	4.9	5318	63	US-08-928-361B-3	Sequence 3, Appli
	32	137.8	4.9	5163	63	US-08-928-361B-4	Sequence 4, Appli
	33	135.2	4.8	7334	26	US-08-928-361-1	Sequence 1, Appli
	34	135.2	4.8	5511	26	US-08-928-361-2	Sequence 2, Appli
	35	135.2	4.8	7334	63	US-08-928-361B-1	Sequence 1, Appli
	36	135.2	4.8	5511	63	US-08-928-361B-2	Sequence 2, Appli
	37	120.6	4.3	565	60	US-60-138-103-14338	Sequence 14338, A
	38	120.2	4.3	1010	60	US-60-138-103-13491	Sequence 13491, A
	39	119.2	4.2	32207	1	PCT-US97-01442-20	Sequence 20, Appl
	40	119.2	4.2	32207	1	PCT-US97-13346-20	Sequence 20, Appl
	41	119.2	4.2	12754	16	US-08-592-963B-58	Sequence 58, Appl
	42	119.2	4.2	12753	16	US-08-592-963B-59	Sequence 59, Appl
	43	119.2	4.2	12752	16	US-08-592-963B-60	Sequence 60, Appl
c	44	119.2	4.2	12754	16	US-08-592-963B-61	Sequence 61, Appl
c	45	119.2	4.2	12753	16	US-08-592-963B-62	Sequence 62, Appl

ALIGNMENTS

RESULT 1

```
US-09-103-429-2
; Sequence 2, Application US/09103429
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,429
;      FILING DATE:  24-JUN-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2821 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      DEVELOPMENTAL STAGE:  larva
;      TISSUE TYPE:  peritrophic membrane
US-09-103-429-2


  Query Match             100.0%;  Score 2821;  DB 33;  Length 2821;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAAAAGAATAACC 13
          |||||||||||||
Db      1 GAAAAGAATAACC 13


RESULT   2
US-09-103-429A-2
; Sequence 2, Application US/09103429A
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,429A
;      FILING DATE:  24-JUN-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2821 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      DEVELOPMENTAL STAGE:  larva
;      TISSUE TYPE:  peritrophic membrane
US-09-103-429A-2


  Query Match             100.0%;  Score 2821;  DB 33;  Length 2821;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAAAAGAATAACC 13
          |||||||||||||
Db      1 GAAAAGAATAACC 13


RESULT   3
US-09-294-663-2
; Sequence 2, Application US/09294663
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga Street
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/294,663
;      FILING DATE:  19-APR-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/103,429
;      FILING DATE:  24-JUN-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
```

```
;      REFERENCE/DOCKET NUMBER:  BTI-39-CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2821 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      DEVELOPMENTAL STAGE:  larva
;      TISSUE TYPE:  peritrophic membrane
US-09-294-663-2


  Query Match           100.0%;  Score 2821;  DB 38;  Length 2821;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2821;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60

Qy      61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120

Qy     121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180

Qy     181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240

Qy     241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300

Qy     301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360

Qy     361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420

Qy     421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480

Qy     481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540

Qy     541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600

Qy     601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660

Qy     661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720

Qy     721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780

Qy     781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840

Qy     841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900

Qy     901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960

Qy     961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020

Qy    1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080

Qy    1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140

Qy    1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200

Qy    1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260

Qy    1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320

Qy    1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380

Qy    1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440

Qy    1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500

Qy    1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560

Qy    1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620

Qy    1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680

Qy    1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740

Qy    1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800

Qy    1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860

Qy    1861 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1920

Qy    1921 ACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 ACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980

Qy    1981 ACCCACCGCAGCCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 ACCCACCGCAGCCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 2040

Qy    2041 CACTGCTGCCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 CACTGCTGCCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100

Qy    2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160

Qy    2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220

Qy    2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280

Qy    2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340

Qy    2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400

Qy    2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460

Qy    2461 TTAATGAAGGTTTTGTTTTATTTTTGAGTTATTATTCCAATGGGCGGGAAAGTCCGCCAT 2520
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 TTAATGAAGGTTTTGTTTTATTTTTGAGTTATTATTCCAATGGGCGGGAAAGTCCGCCAT 2520

Qy    2521 TATTGGGTCTTGCCAGTTTTGAGGAAACCTTTTTTTTTTACTACCAACATTCTTGTGAACC 2580
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 TATTGGGTCTTGCCAGTTTTGAGGAAACCTTTTTTTTTTACTACCAACATTCTTGTGAACC 2580

Qy    2581 CATATTTATTACGTATTAAACATCGTGATTTGAAAAACGTTACATGATTTTTTCATTAAT 2640
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 CATATTTATTACGTATTAAACATCGTGATTTGAAAAACGTTACATGATTTTTTCATTAAT 2640

Qy    2641 TTGAAACAATTTATGTTGTTTTTGTTCTCATTAAATATCAAATATCATTTTCGAAACTGG 2700
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 TTGAAACAATTTATGTTGTTTTTGTTCTCATTAAATATCAAATATCATTTTCGAAACTGG 2700

Qy    2701 CAATTTTGGATTGGAATAATCAACAAATGGTTAAGAAAAAAAACGATTTCTTAAAAATGT 2760
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 CAATTTTGGATTGGAATAATCAACAAATGGTTAAGAAAAAAAACGATTTCTTAAAAATGT 2760

Qy    2761 ATTTATTATAAAATGTGTAAATAAATATACAAATTAGCATTTAAAAAAAAAAAAAAAAAA 2820
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 ATTTATTATAAAATGTGTAAATAAATATACAAATTAGCATTTAAAAAAAAAAAAAAAAAA 2820

Qy    2821 A 2821
           |
Db    2821 A 2821

RESULT   4
```

```
US-09-103-429-1
; Sequence 1, Application US/09103429
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,429
;      FILING DATE:  24-JUN-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2455 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  Peritrophic Membrane
;    IMMEDIATE SOURCE:
;      CLONE:  IIM14
US-09-103-429-1


  Query Match             83.5%;  Score 2354.4;  DB 33;  Length 2455;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 2422;  Conservative    0;  Mismatches    1;  Indels   57;  Gaps    1;

Qy       1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 72

Qy      61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 132

Qy     121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 192

Qy     181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     193 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 252

Qy     241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     253 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 312
```

```
Qy    301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    313 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 372

Qy    361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    373 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 432

Qy    421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 492

Qy    481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    493 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 552

Qy    541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    553 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 612

Qy    601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    613 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 672

Qy    661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    673 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 732

Qy    721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    733 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 792

Qy    781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    793 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 852

Qy    841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    853 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 912

Qy    901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    913 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 972

Qy    961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    973 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1032

Qy   1021 CCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1033 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1092

Qy   1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1093 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1152

Qy   1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1153 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1212

Qy   1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1213 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1272

Qy   1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1273 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1332

Qy   1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1333 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1392

Qy   1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1393 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1452

Qy   1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1453 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1512

Qy   1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1513 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1572

Qy   1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1573 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1632

Qy   1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1633 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1692

Qy   1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1693 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1752

Qy   1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1753 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1812

Qy   1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1813 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1872

Qy   1861 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1873 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1932

Qy   1921 ACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980
          ||||||||||||
Db   1933 ACCACCCACTGCT---------------------------------------------- 1946

Qy   1981 ACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 2040
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1946 ----------GCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 1995

Qy   2041 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1996 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCAACACCACAGCTGC 2055

Qy   2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2056 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2115

Qy   2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2116 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2175

Qy   2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2176 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2235

Qy   2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2236 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2295

Qy   2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2296 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2355

Qy   2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2356 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2415

Qy   2461 TTAATGAAGGTTTTGTTTTA 2480
          ||||||||||||||||||||
```

```
Db   2416 TTAATGAAGGTTTTGTTTTA 2435


RESULT   5
US-09-103-429A-1
; Sequence 1, Application US/09103429A
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,429A
;      FILING DATE:  24-JUN-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2455 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  Peritrophic Membrane
;    IMMEDIATE SOURCE:
;      CLONE:  IIM14
US-09-103-429A-1


  Query Match             83.5%;  Score 2354.4;  DB 33;  Length 2455;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 2422;  Conservative    0;  Mismatches    1;  Indels   57;  Gaps    1;

Qy       1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 72

Qy      61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 132

Qy     121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 192

Qy     181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     193 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 252

Qy     241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     253 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 312

Qy     301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     313 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 372

Qy     361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     373 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 432

Qy     421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     433 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 492

Qy     481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     493 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 552

Qy     541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     553 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 612

Qy     601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     613 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 672

Qy     661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     673 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 732

Qy     721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     733 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 792

Qy     781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     793 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 852

Qy     841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     853 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 912

Qy     901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     913 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 972

Qy     961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     973 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1032

Qy    1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1033 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1092

Qy    1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1093 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1152

Qy    1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1153 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1212

Qy    1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1213 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1272

Qy    1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1273 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1332

Qy    1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380
```

```
Db   1333 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1392

Qy   1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440

Db   1393 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1452

Qy   1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500

Db   1453 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1512

Qy   1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560

Db   1513 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1572

Qy   1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620

Db   1573 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1632

Qy   1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680

Db   1633 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1692

Qy   1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740

Db   1693 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1752

Qy   1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800

Db   1753 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1812

Qy   1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860

Db   1813 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1872

Qy   1861 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1920

Db   1873 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1932

Qy   1921 ACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980

Db   1933 ACCACCCACTGCT---------------------------------------------- 1946

Qy   1981 ACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 2040

Db   1946 ----------GCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 1995

Qy   2041 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100

Db   1996 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCAACACCACAGCTGC 2055

Qy   2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160

Db   2056 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2115

Qy   2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220

Db   2116 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2175

Qy   2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280

Db   2176 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2235

Qy   2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340

Db   2236 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2295

Qy   2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400

Db   2296 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2355

Qy   2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460

Db   2356 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2415

Qy   2461 TTAATGAAGGTTTTGTTTTA 2480

Db   2416 TTAATGAAGGTTTTGTTTTA 2435
```

```
RESULT   6
US-09-294-663-1
; Sequence 1, Application US/09294663
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga Street
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/294,663
;      FILING DATE:  19-APR-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/103,429
;      FILING DATE:  24-JUN-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39-CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2455 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  Peritrophic Membrane
;    IMMEDIATE SOURCE:
;      CLONE:  IIM14
US-09-294-663-1


  Query Match             83.5%;  Score 2354.4;  DB 38;  Length 2455;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 2422;  Conservative    0;  Mismatches    1;  Indels   57;  Gaps    1;

Qy       1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60

Db      13 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 72

Qy      61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120

Db      73 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 132

Qy     121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180
```

```
Db    133 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 192

Qy    181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240

Db    193 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 252

Qy    241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300

Db    253 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 312

Qy    301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360

Db    313 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 372

Qy    361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420

Db    373 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 432

Qy    421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480

Db    433 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 492

Qy    481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540

Db    493 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 552

Qy    541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600

Db    553 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 612

Qy    601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660

Db    613 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 672

Qy    661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720

Db    673 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 732

Qy    721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780

Db    733 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 792

Qy    781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840

Db    793 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 852

Qy    841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900

Db    853 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 912

Qy    901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960

Db    913 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 972

Qy    961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020

Db    973 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1032

Qy   1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080

Db   1033 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1092

Qy   1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140

Db   1093 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1152

Qy   1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200

Db   1153 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1212

Qy   1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260

Db   1213 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1272

Qy   1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320

Db   1273 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1332

Qy   1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380

Db   1333 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1392

Qy   1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440

Db   1393 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1452

Qy   1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500

Db   1453 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1512

Qy   1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560

Db   1513 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1572

Qy   1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620

Db   1573 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1632

Qy   1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680

Db   1633 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1692

Qy   1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740

Db   1693 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1752

Qy   1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800

Db   1753 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1812

Qy   1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860

Db   1813 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1872

Qy   1861 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1920

Db   1873 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1932

Qy   1921 ACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980

Db   1933 ACCACCCACTGCT---------------------------------------------- 1946

Qy   1981 ACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 2040

Db   1946 ----------GCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 1995

Qy   2041 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100

Db   1996 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCAACACCACAGCTGC 2055

Qy   2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160

Db   2056 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2115

Qy   2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220

Db   2116 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2175

Qy   2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280

Db   2176 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2235

Qy   2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340

Db   2236 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2295
```

```
Qy   2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2296 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2355

Qy   2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2356 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2415

Qy   2461 TTAATGAAGGTTTTGTTTTA 2480
          ||||||||||||||||||||
Db   2416 TTAATGAAGGTTTTGTTTTA 2435


RESULT   7
PCT-US99-06742-14
; Sequence 14, Application PC/TUS9906742
; GENERAL INFORMATION:
;  APPLICANT: Ruffner, Duane E.
;  APPLICANT: Pierce, Michael L.
;  APPLICANT: Chen, Zhidong
;  TITLE OF INVENTION: Directed Antisense Libraries
;  FILE REFERENCE: T6678.PCT
;  CURRENT APPLICATION NUMBER: PCT/US99/06742
;  CURRENT FILING DATE: 1999-03-28
;  EARLIER APPLICATION NUMBER: 60/079,792
;  EARLIER FILING DATE: 1998-03-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: Word97
; SEQ ID NO 14
;   LENGTH: 8705
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: pShuttle
PCT-US99-06742-14


  Query Match             5.6%;  Score 158.8;  DB 1;  Length 8705;
  Best Local Similarity   53.3%;  Pred. No. 5.3e-26;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db   7304 cctccacctcctgctcctgccctcctgctcctgccctcctcctgctcctgccctcct 7363

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || |  | |      | ||| |  | || || ||  ||||  | || | ||| ||| |
Db   7364 gcccctcctgctcctgccctcctgccctcctgctcctgccctcctgccctcctgct 7423

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
          | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db   7424 cctgccctcctgccctcctcctgctcctgccctcctgccctcctcctgctcctgcc 7483

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||   | ||  |  ||| | |     | ||| |||||  || || |  | ||||||
Db   7484 cctcctgccctcctgctcctgccctcctgccctcctgctcctgccctcctgcccct 7543

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
          || || ||| ||||      ||   | ||    |  | || ||| | ||| |  || |
Db   7544 cctgctcctgcccctcctgctcctgccctcctgctcctgccctcctgctcctgcccct 7603

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
          |||||||  | ||  | ||  | ||  |  |  || |||||     |         |  ||
Db   7604 cctgccctcctgccctcctcctgctcctgccctcctgctcctgccctcctgcccct 7663

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCCAACACCACA 1912
          ||| | |  |  | ||| |  || | ||  || | ||  | ||||||||     || |
Db   7664 cctgccctcctgctcctgccctcctcctgctcctgccctcctgccctcctgcccct 7723

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
            |||  ||  || || |||||||| || | ||  |  |  ||||     |  | |
Db   7724 cctcctgctcctgccctcctgccctcctcctgctcctgccctcctcctgctcctgcc 7783

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
          || |  | ||  | ||  | ||| || |  | || ||| |  ||    |  | |  ||
Db   7784 cctcctgcccctcctgcccctcctcctgctcctgcccctcctgcccctcctcctgctcct 7843

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
          |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||   | |
Db   7844 gcccctcctcctgctcctgccctcctgccctcctgccctcctcctgctcctgcccct 7903

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
          |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db   7904 cctcctgctcctgccctcctgccctcctgccctcctgccctcctcctgctcctgcc 7963

Qy   2153 GACCCCCCTCTCCC 2166
             || |||   ||
Db   7964 cctcctcctgctcc 7977


RESULT   8
US-08-407-128-15/c
; Sequence 15, Application US/08407128
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/407,128
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  779
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-407-128-15


  Query Match             5.6%;  Score 158.8;  DB 13;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 5.9e-26;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || |  | |      | ||| |  | || || ||  ||||  | || | ||| ||| |
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827
```

```
Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
Db   2526 CCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
Db   2286 CCTCCTCCTGCTCC 2273
```

```
RESULT   9
US-08-409-770-15/c
; Sequence 15, Application US/08409770
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/409,770
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  755.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-409-770-15

  Query Match             5.6%;  Score 158.8;  DB 13;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 5.9e-26;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
Db   2526 CCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCT 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
Db   2286 CCTCCTCCTGCTCC 2273
```

```
RESULT  10
US-08-487-827-15/c
; Sequence 15, Application US/08487827
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J., Lokker, Natalie A., Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
```

```
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,827
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/087783
;      FILING DATE:  13-JULY-93
;      APPLICATION NUMBER:  07/884811
;      FILING DATE:  18-MAY-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/885971
;      FILING DATE:  18-MAY-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gallegos, Thomas R.
;      REGISTRATION NUMBER:  32,692
;      REFERENCE/DOCKET NUMBER:  755,779P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-2614
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-487-827-15


  Query Match             5.6%;  Score 158.8;  DB 15;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 5.9e-26;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || |  | |      | ||| |  | || || ||  ||||  | || | ||| ||| |
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
          | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||   | ||  |  ||| | |     | ||| |||||  || || |  | ||||||
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
          || || ||| ||||      ||   | ||    |  | || ||| | ||| |  || |
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
           |||||||  | ||  | ||  | ||  |  |  || |||||    |       |  ||
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
          |||  |  |   |  | |||  |  ||  |  ||  ||  |  ||  | ||||||||     ||  |
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
             |||  ||  || || |||||||| || |  ||  |  |  ||||     |  | |

Db   2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
          || |  | ||  | ||  | ||| || |  | || ||| |  ||     |  | |   ||
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
          |  || ||  ||||| |  || || | || || ||| ||     | |||||  ||    | |
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
          |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
             || |||   ||
Db   2286 CCTCCTCCTGCTCC 2273


RESULT  11
US-08-487-831-15/c
; Sequence 15, Application US/08487831
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,831
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  779
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-487-831-15


  Query Match             5.6%;  Score 158.8;  DB 15;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 5.9e-26;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || |  | |      | ||| |  | || || ||  ||||  | || | ||| ||| |
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827
```

```
Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
Db   2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
Db   2286 CCTCCTCCTGCTCC 2273
```

```
RESULT  12
US-08-733-369A-104/c
; Sequence 104, Application US/08733369A
;  GENERAL INFORMATION:
;    APPLICANT:  Masucci, Maria G.
;    TITLE OF INVENTION:  GLYCINE-CONTAINING SEQUENCES CONFERRING
;    TITLE OF INVENTION:  INVISIBILITY TO THE IMMUNE SYSTEM.
;    NUMBER OF SEQUENCES:  123
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kathleen M. Williams, Banner & Witcoff, Ltd.
;      STREET:  One Financial Center
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Wordperfect 6.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,369A
;      FILING DATE:  17 October, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/522,995
;      FILING DATE:  01-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/529,190
;      FILING DATE:  15-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 95013249
;      FILING DATE:  10-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB96/00876
;      FILING DATE:  10-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Kathleen M.
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  95-1391-D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-345-9100
;      TELEFAX:  617-345-9111
;  INFORMATION FOR SEQ ID NO:  104:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  803 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-733-369A-104
```

```
  Query Match             5.6%;  Score 158.8;  DB 18;  Length 803;
  Best Local Similarity   53.3%;  Pred. No. 1.3e-26;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
Db    745 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 686

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
Db    685 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 626

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db    625 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 566

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db    565 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 506

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db    505 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 446

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
Db    445 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 386

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
Db    385 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 326

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
Db    325 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 266

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
Db    265 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 206

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
Db    205 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 146

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
Db    145 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 86

Qy   2153 GACCCCCCTCTCCC 2166
Db     85 CCTCCTCCTGCTCC 72
```

```
RESULT  13
US-08-733-369A-105/c
; Sequence 105, Application US/08733369A
;  GENERAL INFORMATION:
;    APPLICANT:  Masucci, Maria G.
;    TITLE OF INVENTION:  GLYCINE-CONTAINING SEQUENCES CONFERRING
;    TITLE OF INVENTION:  INVISIBILITY TO THE IMMUNE SYSTEM.
;    NUMBER OF SEQUENCES:  123
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kathleen M. Williams, Banner & Witcoff, Ltd.
;      STREET:  One Financial Center
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Wordperfect 6.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,369A
;      FILING DATE:  17 October, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/522,995
;      FILING DATE:  01-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/529,190
;      FILING DATE:  15-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 95013249
;      FILING DATE:  10-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB96/00876
;      FILING DATE:  10-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Kathleen M.
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  95-1391-D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-345-9100
;      TELEFAX:  617-345-9111
;  INFORMATION FOR SEQ ID NO:  105:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  799 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-733-369A-105


  Query Match             5.6%;  Score 158.8;  DB 18;  Length 799;
  Best Local Similarity   53.3%;  Pred. No. 1.3e-26;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db    741 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 682

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || |  | |       | ||| |  | || || ||  ||||  | || | ||| ||| |
Db    681 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 622

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
          | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db    621 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 562

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||  | ||  |  ||| | |     | ||| ||||| || || |  | ||||||
Db    561 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 502

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
          || || ||| ||||      ||   | ||    |  | || ||| | ||| |  || |
Db    501 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 442

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
          |||||||  | ||  | ||  | ||  |  |  || |||||    |        |  ||
Db    441 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 382

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
          |||  |  |   |  ||| |  || | ||  | | ||  | ||||||||     || |
Db    381 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 322

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
             |||  ||  || || |||||||| || | ||  |  |  ||||     |  | |
Db    321 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 262

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
          || |  | ||  | ||  | ||| || |  | || ||| |  ||    |  | |  ||
Db    261 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 202

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
          |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||   | |
Db    201 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 142

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
          |  |||| || |   || |  ||    |  | || |  | |||     || |     |
Db    141 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 82

Qy   2153 GACCCCCCTCTCCC 2166
             || |||   ||
Db     81 CCTCCTCCTGCTCC 68


RESULT  14
US-08-733-369A-112/c
; Sequence 112, Application US/08733369A
;  GENERAL INFORMATION:
;    APPLICANT:  Masucci, Maria G.
;    TITLE OF INVENTION:  GLYCINE-CONTAINING SEQUENCES CONFERRING
;    TITLE OF INVENTION:  INVISIBILITY TO THE IMMUNE SYSTEM.
;    NUMBER OF SEQUENCES:  123
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kathleen M. Williams, Banner & Witcoff, Ltd.
;      STREET:  One Financial Center
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Wordperfect 6.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,369A
;      FILING DATE:  17 October, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/522,995
;      FILING DATE:  01-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/529,190
;      FILING DATE:  15-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 95013249
;      FILING DATE:  10-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB96/00876
;      FILING DATE:  10-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Kathleen M.
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  95-1391-D
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  617-345-9100
;      TELEFAX:  617-345-9111
;  INFORMATION FOR SEQ ID NO:  112:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  798 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-733-369A-112


  Query Match             5.6%;  Score 158.8;  DB 18;  Length 798;
  Best Local Similarity   53.3%;  Pred. No. 1.3e-26;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db    740 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 681

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || |  | |       | ||| |  | || || ||  ||||  | || | ||| ||| |
Db    680 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 621

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
          | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db    620 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 561

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||   | ||  |  ||| | |     | ||| ||||| || || |  | ||||||
Db    560 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 501

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
          || || ||| ||||      ||   | ||    |  | || ||| | ||| |  || |  ||
Db    500 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 441

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
          |||||||  | ||  | ||  | ||  |  |  || |||||    |       |  ||
Db    440 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 381

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
          |||  |  |  |  | |||  |  || | ||  || |  ||  | ||||||||     || |
Db    380 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 321

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
              |||  ||  || || |||||||| || |  ||  |  |  ||||     |  | |
Db    320 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 261

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
          || |  | ||  | ||  | ||| || |  | || ||| |  ||    |  | |   ||
Db    260 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 201

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
          |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||   | |
Db    200 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 141

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
          |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db    140 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 81

Qy   2153 GACCCCCCTCTCCC 2166
          || |||   ||
Db     80 CCTCCTCCTGCTCC 67


RESULT  15
US-08-910-647-1/c
; Sequence 1, Application US/08910647
;  GENERAL INFORMATION:
;    APPLICANT:  Zuckermann et al.
;    TITLE OF INVENTION:  Compositions and Methods for
;    TITLE OF INVENTION:  Polynucleotide Delivery
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,647
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fujita, Sharon M.
;      REGISTRATION NUMBER:  38,459
;      REFERENCE/DOCKET NUMBER:  1218.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 923-2706
;      TELEFAX:  (510) 655-3542
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9600 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-910-647-1


  Query Match             5.6%;  Score 158.8;  DB 25;  Length 9600;
  Best Local Similarity   53.3%;  Pred. No. 5.6e-26;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db   1412 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 1353

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || |  | |       | ||| |  | || || ||  ||||  | || | ||| ||| |
Db   1352 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 1293

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
          | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db   1292 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 1233

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||   | ||  |  ||| | |     | ||| ||||| || || |  | ||||||
Db   1232 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 1173

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
          || || ||| ||||      ||   | ||    |  | || ||| | ||| |  || |  ||
Db   1172 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 1113

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
          |||||||  | ||  | ||  | ||  |  |  || |||||    |       |  ||
Db   1112 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 1053

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
          |||  |  |  |  | |||  |  || | ||  || |  ||  | ||||||||     || |
Db   1052 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 993

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
              |||  ||  || || |||||||| || |  ||  |  |  ||||     |  | |
Db    992 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 933

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
          || |  | ||  | ||  | ||| || |  | || ||| |  ||    |  | |   ||
Db    932 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 873
```

```
Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
          |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||   | |
Db    872 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 813

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
           |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db    812 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 753

Qy   2153 GACCCCCCTCTCCC 2166
             || |||   ||
Db    752 CCTCCTCCTGCTCC 739


Search completed: July 18, 1999, 02:09:55
Job time: 3109 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 07:53:30 ; Search time 72.5 Seconds
(without alignments)
3567.166 Million cell updates/sec

Title: US-09-294-663-2
Perfect score: 2821
Sequence: 1 GAAAAGAATAACCAGCGAAC.....TAAAAAAAAAAAAAAAAAAA 2821

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	158.8	5.6	10596	1	US-07-884-811-15	Sequence 15, Appl
c	2	158.8	5.6	10596	1	US-07-885-971-15	Sequence 15, Appl
c	3	158.8	5.6	10596	1	US-08-087-783A-15	Sequence 15, Appl
c	4	158.8	5.6	10596	1	US-08-194-088B-15	Sequence 15, Appl
c	5	158.8	5.6	10596	4	US-08-194-087-15	Sequence 15, Appl
c	6	158.8	5.6	10596	5	PCT-US93-04648-15	Sequence 15, Appl
c	7	120.2	4.3	2793	3	US-08-209-747-1	Sequence 1, Appli
c	8	120.2	4.3	2793	3	US-08-458-298-1	Sequence 1, Appli
	9	119.2	4.2	32207	4	US-08-770-379-20	Sequence 20, Appl
c	10	110.8	3.9	6530	5	PCT-US93-03993-1	Sequence 1, Appli
c	11	105	3.7	9551	1	US-08-056-200-93	Sequence 93, Appl
	12	101	3.6	1984	1	US-07-885-970A-25	Sequence 25, Appl
	13	101	3.6	1985	1	US-08-298-687A-25	Sequence 25, Appl
	14	101	3.6	1985	1	US-08-298-829-25	Sequence 25, Appl
	15	100	3.5	913	1	US-08-217-327-3	Sequence 3, Appli
	16	100	3.5	913	1	US-07-885-970A-3	Sequence 3, Appli
	17	100	3.5	913	1	US-08-298-687A-3	Sequence 3, Appli
	18	100	3.5	913	1	US-08-530-797-2	Sequence 2, Appli
	19	100	3.5	913	1	US-08-298-829-3	Sequence 3, Appli
c	20	90.4	3.2	44377	4	US-08-804-227C-7	Sequence 7, Appli
c	21	90.4	3.2	13987	4	US-08-804-227C-13	Sequence 13, Appl
c	22	88.8	3.1	1505	1	US-07-915-246-1	Sequence 1, Appli
	23	84.4	3.0	12001	3	US-08-458-568A-11	Sequence 11, Appl
	24	80.4	2.9	7218	2	US-08-232-463-14	Sequence 14, Appl
c	25	80	2.8	43280	4	US-08-804-227C-1	Sequence 1, Appli
c	26	79.8	2.8	2338	2	US-08-425-069-1	Sequence 1, Appli
c	27	77.2	2.7	10747	5	PCT-US93-03985-1	Sequence 1, Appli
c	28	77.2	2.7	24979	5	PCT-US93-03985-3	Sequence 3, Appli
	29	74.2	2.6	3833	3	US-08-917-320-18	Sequence 18, Appl
	30	74.2	2.6	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
c	31	73.2	2.6	2242	2	US-08-641-627A-37	Sequence 37, Appl
c	32	72.4	2.6	2427	4	US-08-678-039A-39	Sequence 39, Appl
c	33	72.2	2.6	32207	4	US-08-770-379-20	Sequence 20, Appl
	34	71	2.5	31571	2	US-08-323-443B-1	Sequence 1, Appli
	35	64.8	2.3	3468	1	US-07-951-715A-2	Sequence 2, Appli
	36	64.8	2.3	786	3	US-08-403-379A-2	Sequence 2, Appli
	37	64.8	2.3	3468	4	US-08-459-448A-2	Sequence 2, Appli
	38	64.6	2.3	2093	1	US-08-287-001A-1	Sequence 1, Appli
	39	64.6	2.3	2093	5	PCT-US95-09941-1	Sequence 1, Appli
	40	63.6	2.3	3624	1	US-07-951-715A-6	Sequence 6, Appli
	41	63.6	2.3	3624	4	US-08-459-448A-6	Sequence 6, Appli
c	42	63.4	2.2	234	3	US-08-469-802B-3	Sequence 3, Appli
c	43	63.4	2.2	234	4	US-08-267-803B-3	Sequence 3, Appli
	44	63.2	2.2	2032	5	PCT-US95-07721-5	Sequence 5, Appli
	45	62.6	2.2	2753	4	US-07-952-853-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT   1
US-07-884-811-15/c
; Sequence 15, Application US/07884811
; Patent No. 5316921
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/884,811
;      FILING DATE:  19920518
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  755.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-884-811-15


  Query Match             5.6%;  Score 158.8;  DB 1;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 1.1e-24;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy    1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
           ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db    2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy    1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
           .|| |  | |       | ||| |  | || || ||  ||||  | || | ||| ||| |
Db    2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy    1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
           | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db    2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767
```

```
Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
Db   2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
Db   2286 CCTCCTCCTGCTCC 2273
```

RESULT 2
US-07-885-971-15/c
; Sequence 15, Application US/07885971
; Patent No. 5328837
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,971
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-885-971-15

Query Match 5.6%; Score 158.8; DB 1; Length 10596;
Best Local Similarity 53.3%; Pred. No. 1.1e-24;
Matches 359; Conservative 0; Mismatches 312; Indels 3; Gaps 1;

```
Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
Db   2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
Db   2286 CCTCCTCCTGCTCC 2273
```

RESULT 3
US-08-087-783A-15/c
; Sequence 15, Application US/08087783A
; Patent No. 5547856
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco

```
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WinPatin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/087,783A
;       FILING DATE:  13-Jul-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/884811
;       FILING DATE:  18-MAY-92
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/885971
;       FILING DATE:  18-MAY-92
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Marschang, Diane L.
;       REGISTRATION NUMBER:  35,600
;       REFERENCE/DOCKET NUMBER:  P0755779P1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-5416
;       TELEFAX:  415/952-9881
;       TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10596 base pairs
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-087-783A-15


  Query Match             5.6%;  Score 158.8;  DB 1;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 1.1e-24;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          || || || || |  | ||  |  | ||  |  ||||||  |  |||| ||  |  | |
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || |  | |      | ||| |  | || || ||  ||||  | || | ||| ||| |
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
          | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||   | ||  |  ||| | |      | ||| ||||| || || |  | ||||||
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
          || || ||| ||||      ||   | ||     |  | || ||| | ||| |  || |
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
          |||||||  | ||  | ||  | ||  |  |  || |||||    |          |  ||
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
          ||| | |  |  | ||| |  || | ||  || | ||  | ||||||||     || |
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
            |||  ||  || || |||||||| || | ||  |  |  ||||     |  | |
Db   2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
          || |  | ||  | ||  | ||| || |  | || ||| |  ||    |  | |   ||

Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
          |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||   | |
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
            |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
             || |||   ||
Db   2286 CCTCCTCCTGCTCC 2273


RESULT   4
US-08-194-088B-15/c
; Sequence 15, Application US/08194088B
; Patent No. 5580963
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,088B
;      FILING DATE:  09-FEB-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/884811
;      FILING DATE:  18-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gallegos, R. Thomas
;      REGISTRATION NUMBER:  32,692
;      REFERENCE/DOCKET NUMBER:  755D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-2614
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-194-088B-15


  Query Match             5.6%;  Score 158.8;  DB 1;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 1.1e-24;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          || || || || |  | ||  |  | ||  |  ||||||  |  |||| ||  |  | |
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || |  | |      | ||| |  | || || ||  ||||  | || | ||| ||| |
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
          | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
```

```
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||   | ||  |  ||| | |     | ||| ||||| || || |  | ||||||
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
             || || ||| ||||       ||   | ||     |  | || ||| | ||| |  || |
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
             |||||||  | ||  | ||  | ||  |  |  || |||||     |         |  ||
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
             ||| | |  |  | ||| |  || | ||  || | ||  | ||||||||     || |
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
             |||  ||  || || ||||||||| || | ||  |  |  ||||     |  | |
Db   2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
             || |  | ||  | ||  | ||| || |  | || ||| |  ||     |  | |   ||
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
             |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||    | |
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
             |  |||| || |   || |  ||     |  | || |  | |||     || |     |
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
             || |||   ||
Db   2286 CCTCCTCCTGCTCC 2273


RESULT   5
US-08-194-087-15/c
; Sequence 15, Application US/08194087
; Patent No. 5879910
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,087
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  779
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-194-087-15


  Query Match             5.6%;  Score 158.8;  DB 4;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 1.1e-24;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
             ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  | |
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
             || |  | |       | ||| |  | || || ||  ||||  | || | ||| ||| |
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
             | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||   | ||  |  ||| | |     | ||| ||||| || || |  | ||||||
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
             || || ||| ||||       ||   | ||     |  | || ||| | ||| |  || |
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
             |||||||  | ||  | ||  | ||  |  |  || |||||     |         |  ||
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
             ||| | |  |  | ||| |  || | ||  || | ||  | ||||||||     || |
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
             |||  ||  || || ||||||||| || | ||  |  |  ||||     |  | |
Db   2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
             || |  | ||  | ||  | ||| || |  | || ||| |  ||     |  | |   ||
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
             |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||    | |
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
             |  |||| || |   || |  ||     |  | || |  | |||     || |     |
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
             || |||   ||
Db   2286 CCTCCTCCTGCTCC 2273


RESULT   6
PCT-US93-04648-15/c
; Sequence 15, Application PC/TUS9304648
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc., Godowski, Paul J., Lokker, Natalie A., Mark, Melani
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
```

```
;       CITY:  South San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  patin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/04648
;       FILING DATE:  19930517
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/884811
;       FILING DATE:  18-MAY-92
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/885971
;       FILING DATE:  18-MAY-92
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Dreger, Ginger R.
;       REGISTRATION NUMBER:  33,055
;       REFERENCE/DOCKET NUMBER:  755,779P1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415/225-3216
;       TELEFAX:  415/952-9881
;       TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10596 bases
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
PCT-US93-04648-15


  Query Match             5.6%;  Score 158.8;  DB 5;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 1.1e-24;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
            || |  | |      | ||| |  | || || ||  ||||  | || | ||| ||| |
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
          | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || ||||| 
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||   | ||  |  ||| | |     | ||| ||||| || || |  | ||||||
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
          || || ||| ||||      ||   | ||    |  | || ||| | ||| |  || |
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
          |||||||  | ||  | ||  | ||  |  |  || |||||    |        |  ||
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
          |||  |  |   |  | ||| |  || | ||  || | ||  | |||||||||     || |
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
              |||  ||  || || |||||||| || | ||  |  |  ||||     |  | |
Db   2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
          || |  | ||  | ||  | ||| || |  | || ||| |  ||    |  | |  ||
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
          |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||   | ||
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
            |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
             || |||   ||
Db   2286 CCTCCTCCTGCTCC 2273


RESULT   7
US-08-209-747-1/c
; Sequence 1, Application US/08209747
; Patent No. 5733771
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/209,747
;      FILING DATE:  14-MAR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-209-747-1


  Query Match             4.3%;  Score 120.2;  DB 3;  Length 2793;
  Best Local Similarity   50.1%;  Pred. No. 8.9e-17;
  Matches  333;  Conservative    0;  Mismatches  323;  Indels    9;  Gaps    1;

Qy   1482 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1541
```

```
Db   2120 CAGCGGCTCCAGCTCCTGCAGCGGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGC 2061

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
          | ||   | | |||||  |  | |  |  |||| ||  |  | || ||| |     | |
Db   2060 CTCCGGATCCTGCACCAGCAGCTGCAGCAGCACCAGCTCCTGCTCCGGCACCGTATCCAC 2001

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
          |    |||||     |  ||  |||| |||| || ||  | || || ||  |  |||| |
Db   2000 CTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCAC 1941

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
          |  |  || |||| |    | | | |        | |||  || ||  || |||||||
Db   1940 CAGCGGCTCCTGCACGACCACCTCTACTGTAACCCGCAGCGCCTCCAGCTCCTGCTCCAG 1881

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
          | || ||| |  |  | ||| |  ||      ||      |   |  | || ||    |
Db   1880 CAGCTGCCGCAGCACCAGCTCTTGCTCCGGCACCGTATCCACCCTGACCACCGTAGCCTC 1821

Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
          | || ||  |  |||| ||  | || || ||  |  |||| ||| |  | |||    |
Db   1820 CAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCACCTGCAGCAGCTCCAGCTC 1761

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
           |   ||| |  || |  | | |||| |  || |||| ||  |  | |||||| | |||
Db   1760 CAGCACCAGCACCTCTACCGTAACCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCG 1701

Qy   1902 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCA---------C 1952
          |     |  | |  | | ||||    |  ||   ||  |     |||||          |
Db   1700 CACCAGCTCCTGCTCCGGCACCATAACCGCCTTGTCCACCATATCCAGCACCACCGGTGC 1641

Qy   1953 CTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTG 2012
          |||| ||   | |  ||| | |  || | ||  |  |||| |||    |   | ||||  |
Db   1640 CTGCACCTGCCCCAGCAGCGGCACCAGCTCCAGCACCAGCACCTCTACCGTAACCACCAG 1581

Qy   2013 CCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCG 2072
          | ||     |  | ||  | | | |  | | ||  |  |  ||||  |  | ||
Db   1580 CGCCTCCGGCTCCTGTTGCAGCAGCTGCAGCAGCACCAGCTCCTGCTCCGGCACCGTATC 1521

Qy   2073 TCGCACCTGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCA 2132
            |     ||||  |  |  ||||| |  ||    |  || |    || |||||  |
Db   1520 CACCTTGACCACCGTAGCCTCCAGCTCCTGCGCCTGCAGCAGCTCCAGCTCCAGCACCAG 1461

Qy   2133 CACCT 2137
          |||||
Db   1460 CACCT 1456


RESULT   8
US-08-458-298-1/c
; Sequence 1, Application US/08458298
; Patent No. 5756677
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,298
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,747
;      FILING DATE:  14-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-458-298-1


  Query Match             4.3%;  Score 120.2;  DB 3;  Length 2793;
  Best Local Similarity   50.1%;  Pred. No. 8.9e-17;
  Matches  333;  Conservative    0;  Mismatches  323;  Indels    9;  Gaps    1;

Qy   1482 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1541
          || || || | || ||  |||| |||||| |  | || ||      |    ||| | ||
Db   2120 CAGCGGCTCCAGCTCCTGCAGCGGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGC 2061

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
          | ||   | | |||||  |  | |  |  |||| ||  |  | || ||| |     | |
Db   2060 CTCCGGATCCTGCACCAGCAGCTGCAGCAGCACCAGCTCCTGCTCCGGCACCGTATCCAC 2001

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
          |    |||||     |  ||  |||| |||| || ||  | || || ||  |  |||| |
Db   2000 CTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCAC 1941

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
          |  |  || |||| |    | | | |        | |||  || ||  || |||||||
Db   1940 CAGCGGCTCCTGCACGACCACCTCTACTGTAACCCGCAGCGCCTCCAGCTCCTGCTCCAG 1881

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
          | || ||| |  |  | ||| |  ||      ||      |   |  | || ||    |
Db   1880 CAGCTGCCGCAGCACCAGCTCTTGCTCCGGCACCGTATCCACCCTGACCACCGTAGCCTC 1821

Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
          | || ||  |  |||| ||  | || || ||  |  |||| ||| |  | |||    |
Db   1820 CAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCACCTGCAGCAGCTCCAGCTC 1761

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
           |   ||| |  || |  | | |||| |  || |||| ||  |  | |||||| | |||
Db   1760 CAGCACCAGCACCTCTACCGTAACCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCG 1701

Qy   1902 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTACCGCAGCA---------C 1952
          |     |  | |  | | ||||    |  ||   ||  |     |||||          |
Db   1700 CACCAGCTCCTGCTCCGGCACCATAACCGCCTTGTCCACCATATCCAGCACCACCGGTGC 1641

Qy   1953 CTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTG 2012
          |||| ||   | |  ||| | |  || | ||  |  |||| |||    |   | ||||  |
Db   1640 CTGCACCTGCCCCAGCAGCGGCACCAGCTCCAGCACCAGCACCTCTACCGTAACCACCAG 1581

Qy   2013 CCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCG 2072
          | ||     |  | ||  | | | |  | | ||  |  |  ||||  |  | ||
Db   1580 CGCCTCCGGCTCCTGTTGCAGCAGCTGCAGCAGCACCAGCTCCTGCTCCGGCACCGTATC 1521
```

```
Qy   2073 TCGCACCTGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCA 2132
Db   1520 CACCTTGACCACCGTAGCCTCCAGCTCCTGCGCCTGCAGCAGCTCCAGCTCCAGCACCAG 1461

Qy   2133 CACCT 2137
Db   1460 CACCT 1456


RESULT   9
US-08-770-379-20
; Sequence 20, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-20


  Query Match             4.2%;  Score 119.2;  DB 4;  Length 32207;
  Best Local Similarity   48.8%;  Pred. No. 3.9e-16;
  Matches  322;  Conservative    0;  Mismatches  338;  Indels    0;  Gaps    0;

Qy   1488 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1547
Db  19231 CTGCTCCTGCTCTTCCACCTCCTCTAACTCCTGCTGCTCCTGCTCTTCCACCTCCTCTAA 19290

Qy   1548 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1607
Db  19291 CTCCTGCTCTTCCTGCTCTTCCACCTCCTCTAACTCCTGCTCTTCCTGCTCTTCCACCTC 19350

Qy   1608 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1667
Db  19351 CTCTAACTCCTGCTCCTCCTGCTCCTCCTGCTCCTGCTCTTGCTCCTCCACCTCCTCTAA 19410

Qy   1668 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1727
Db  19411 TTCCTGCTCTTCCTGCTCCTGCTCTTGCTCTTCCACCTCCTGCTCTTGCTCTTCCACCTC 19470

Qy   1728 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCC 1787
Db  19471 CTGCTCCTCTAACTCCTGCTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTCTAACTCCTG 19530

Qy   1788 CCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCC 1847
Db  19531 CTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTCTAACTC 19590

Qy   1848 CAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACA 1907
Db  19591 CTGCTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTCTAA 19650

Qy   1908 CCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCA 1967
Db  19651 CTCCTGCTCCTGATCCTCTAACTCCTGCTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTC 19710

Qy   1968 CAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAG 2027
Db  19711 CTGCTGCTCCTGCTCCTCCTGCTGCTCCTGTTCATCCTGCTGCTGCTGCTCATCCTGCTG 19770

Qy   2028 TCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCA 2087
Db  19771 CTGCTGCTCATCCTGCTGCTGCTGCTCATCCTGCTGCTGCTGCTCATCCTGCTGCTGCTG 19830

Qy   2088 ACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATG 2147
Db  19831 CTCATCCTGCTGCTGCTCATCCTGCTGCTCCTGCTCATCCTGCTGCTCCTGCTCATCCTG 19890


RESULT  10
PCT-US93-03993-1/c
; Sequence 1, Application PC/TUS9303993
;  GENERAL INFORMATION:
;    APPLICANT:  Roop, Dennis R.
;    APPLICANT:  Rothnagel, Joseph A.
;    APPLICANT:  Greenhalgh, David A.
;    TITLE OF INVENTION:  CONSTITUTIVE AND INDUCIBLE EPIDERMAL
;    TITLE OF INVENTION:  VECTOR SYSTEMS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fulbright & Jaworski
;      STREET:  1301 McKinney, Suite 5100
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77010-3095
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03993
;      FILING DATE:  19930428
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul, Thomas D.
;      REGISTRATION NUMBER:  32,714
;      REFERENCE/DOCKET NUMBER:  D-5405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713/651-5325
;      TELEFAX:  713/651-5246
;      TELEX:  762829
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6530 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
```

```
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US93-03993-1


  Query Match             3.9%;  Score 110.8;  DB 5;  Length 6530;
  Best Local Similarity   48.4%;  Pred. No. 1.2e-14;
  Matches  307;  Conservative    0;  Mismatches  327;  Indels    0;  Gaps    0;

Qy   1500 CAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCT 1559
          | || ||  | || |  |   | ||  | ||  |    || ||| | ||-|    | ||
Db   3490 CCGCCGCCGCCACCGGAGGAACCACCTCCGCAGCTCGTGCCGCCGCCACCGGAATAGCCG 3431

Qy   1560 CCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCG 1619
          || |||  | | || ||| | |  |  || |    | |||| | | | || ||| | |||
Db   3430 CCGCCGCCGCCGCAGCTGGATCCACCGCCACCTGAGTAGCCGCCGCCACAGCTAGAGCCG 3371

Qy   1620 CAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTG 1679
          |   | || |  | ||| ||       ||| || ||||||||    | || || ||| |
Db   3370 CCGCTGCTTCCACCGCCGCCGGAATAGCCGCCTCCGCAGCTGGAGCCACCGCCGCCACCA 3311

Qy   1680 AATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTG 1739
           |   |  || |   | |    |  | | || |     | || ||| | |     |
Db   3310 GAGTACTTGCCGCCTCCGGAACCGCCGCCGCAGCCACCGGAGCTGCCGCCTCCAGAGGAG 3251

Qy   1740 CTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCC 1799
          |   | | | |    |   | |   |  ||| ||  |  | ||||   || ||||  |||
Db   3250 CCTCCGCAGTAGGAGCCGCCGCCTCCTGATCCGCCTCCGCAGCTGGAGCCGCCGCCGCCG 3191

Qy   1800 CCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCA 1859
          || |||  |   || |||| ||  |||   | || | | |    | | |         |
Db   3190 CCCCCGGAGGAGCCGCCGCCGCAGCTAGAGCCGCCGCCGCCACCACCTCCGGAGTACTTG 3131

Qy   1860 CATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTG 1919
            | ||  | |  || ||| || |  || ||||    ||| ||  |     ||   | |
Db   3130 ACGCCCCCACCAGAGCCGCCACCGCCTCCACAGCTGGAGCCACCCCCGGAGTAGCTGCCG 3071

Qy   1920 TACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCAC 1979
            |  |  |||  ||| |  |  || | | |   ||  ||  | |  | |   | ||
Db   3070 CCGCCGCAGCTGGAGCCGCCGCCGCCTCCGGAGTAGCCACCTCCGCAGCTGGAGCCGCCG 3011

Qy   1980 CACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCAC 2039
          |  || ||  ||   | ||  |||||||    ||| ||| | || | |  |     |  |
Db   3010 CCGCCCCCAGAGTAGCCACCGCCGCAGCTAGAGCCACCACCGCCTCCGGAGTAGCCGCCG 2951

Qy   2040 CCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTG 2099
          || | |||    |||| || ||  ||       | |   |  | ||   || |  |
Db   2950 CCGCAGCTAGAGCCACCGCCGCCTCCGGAGTACTTGACGCCCCCACCGGAGCCGCCTCCA 2891

Qy   2100 CCCCGGTAACTACAACCAGCGCACCAGCTACCAC 2133
          || | | | ||  ||||| | |   ||  |||||
Db   2890 CCGCCGCAGCTGGAACCACCTCCATAGGAACCAC 2857


RESULT  11
US-08-056-200-93/c
; Sequence 93, Application US/08056200
; Patent No. 5616500
;  GENERAL INFORMATION:
;    APPLICANT:  Steinert, Peter M.
;    APPLICANT:  Lee, Seung-Chul
;    APPLICANT:  Kim, In-Gyu
;    APPLICANT:  Chung, Soo-Il
;    APPLICANT:  Park, Sang-Chul
;    TITLE OF INVENTION:  Trichohyalin and Transglutaminase-3 and
;    TITLE OF INVENTION:  Mehods of Using Same
;    NUMBER OF SEQUENCES:  117
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/056,200
;      FILING DATE:  30-APR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fedrick, Michael F.
;      REGISTRATION NUMBER:  36,799
;      REFERENCE/DOCKET NUMBER:  NIH054.001A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (714) 760-0404
;      TELEFAX:  (714) 760-9502
;  INFORMATION FOR SEQ ID NO:  93:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9551 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1507..1644
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1645..2511
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2512..8070
US-08-056-200-93


  Query Match             3.7%;  Score 105;  DB 1;  Length 9551;
  Best Local Similarity   49.3%;  Pred. No. 2.4e-13;
  Matches  310;  Conservative    0;  Mismatches  310;  Indels    9;  Gaps    1;

Qy   1461 GCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCC 1520
          ||| ||| | ||| |    | |  |  ||||  | |  |||| ||| |    ||  | |
Db   3810 GCGCTTCAGCCAATCGCGCCTCTCCTCCTGCTCGCGCTTCAGCCGCTGCTCGCGCCTCTC 3751

Qy   1521 CAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAA 1580
          |  |   |||  |  |   ||||  |  || |  |  |   | ||   |   |  ||
Db   3750 CTGCTCGTGCTTCTGCTCGTGCCTCTCCTCCTCCTGCTCGCGCCTCAGCTGCTGCTCGCG 3691

Qy   1581 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1640
          |  || | ||  ||   ||| |||   |  ||     || ||  | || |||  ||| |
Db   3690 CCTCAGCTGCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCAG 3631

Qy   1641 CCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1700
          | ||||| | | || ||||| || ||   ||  |  |||   | |   || | | |  | |
Db   3630 CTGCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCAGCTGCTG 3571

Qy   1701 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1760
            |   |  | ||||||  |   |||  |  || || ||   | | ||      ||
Db   3570 CTCGCGCCTCAGCTGCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCTCCTCCTCCTGCTC 3511

Qy   1761 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1820
            | | || | ||  ||  |||  | ||  |  |||| | | || | ||  |  |    |
Db   3510 GCGCCTCAGCTGCTGCTCGCGCCTCTCCTCCTCCTGCTCGCGCCTCAGCTGCTGCTCGCG 3451

Qy   1821 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACT------- 1874
          || || || || |||  |  |       |  || ||  | | || || ||| ||
Db   3450 CCTCTCCTCCTCCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCTCCTCCTGCTCGCGCCT 3391

Qy   1874 --GCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTG 1931
            |||||  |     || |  | |||||  |   | || |     |||  |   || ||
Db   3390 CAGCTGCTGCTCGCGCCTCTCCTCCTGCTGCTCGCGCCTCTCCTGCTGCTCGCGCCTCTC 3331
```

```
Qy   1932 CTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAG 1991
          || |  || ||  |||  | ||| ||||      | |  |   |||  | | | || |||
Db   3330 CTCCTGCTGCTCGCGCCTCTCCTCCTCCTGCTTGCGCCTTAGTTGCTGCTCGCGCCTCAG 3271

Qy   1992 CCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCC 2051
          ||  | || | || || | | | ||   | |  |   | | |  |   || || | ||
Db   3270 CCTTTGCTGCTGCTGCTCTTCCTCCTGGCGCTCCCTCCTCAGCTCTTGCCGCTCCAGCTT 3211

Qy   2052 CCACTGCAGCTCCCCCTACCGTCGCACCT 2080
          ||   || ||||  | | |    | | ||
Db   3210 CCGTAGCTGCTCTTCTTCCTCCTGGAGCT 3182


RESULT  12
US-07-885-970A-25
; Sequence 25, Application US/07885970A
; Patent No. 5495070
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,970A
;      FILING DATE:  19920518
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1984 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium barbadense
;      STRAIN:  Sea Island
;    IMMEDIATE SOURCE:
;      LIBRARY:  EMBL SI
;      CLONE:  SIH6
US-07-885-970A-25


  Query Match             3.6%;  Score 101;  DB 1;  Length 1984;
  Best Local Similarity   51.4%;  Pred. No. 8.6e-13;
  Matches  233;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;

Qy   1581 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1640
          |||||||| || |  | ||  |    || ||| | ||  |      |  | ||  | || |
Db    424 CGCCACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 483

Qy   1641 CCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1700
          ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    484 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 543

Qy   1701 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1760
             | ||  |  | | | | ||  |  | ||  | | || ||   |||| ||  |||
Db    544 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 603

Qy   1761 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1820
          |||   |  |  | ||| |  || | ||  | ||| | ||  |  ||||| |  |  ||
Db    604 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 663

Qy   1821 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1880
          |  || || |  ||  | |       ||| || |   | || | ||| | ||| || | |
Db    664 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 723

Qy   1881 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1940
          |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    724 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 783

Qy   1941 CTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTA 2000
          |     | ||||     |     |  |||   |  |  |     || |||| |||||
Db    784 CAGTACAAACACCATTGACATCGCCACCAGCTCCGCCGACCGAGGCCCCAGCACCTACCC 843

Qy   2001 CCGCAGCACCTGCCCCCAACACCACAGTCACTG 2033
          ||  ||  || | ||     |  | |  ||||
Db    844 TCGGGGCTGCTACGCCAGGTCCAGCTGGAACTG 876


RESULT  13
US-08-298-687A-25
; Sequence 25, Application US/08298687A
; Patent No. 5521078
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    TITLE OF INVENTION:  PLANTS FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/298,687A
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1985 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Gossypium barbadense
;       STRAIN:  Sea Island
;     IMMEDIATE SOURCE:
;       LIBRARY:  EMBL-SI
;       CLONE:  SIH6
US-08-298-687A-25


  Query Match             3.6%;  Score 101;  DB 1;  Length 1985;
  Best Local Similarity   51.4%;  Pred. No. 8.6e-13;
  Matches  233;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;

Qy   1581 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1640
          |||||||| || |  | ||  |    || ||| | ||  |      |  | ||  | || |
Db    424 CGCCACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 483

Qy   1641 CCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1700
          ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    484 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 543

Qy   1701 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1760
            | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||
Db    544 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 603

Qy   1761 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1820
          |||   |  |  | ||| |  || | ||  | ||| | ||  |  ||||| |  |  ||
Db    604 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 663

Qy   1821 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1880
          |  || || |  ||  | |       ||| || |   | || | ||| | ||| || | |
Db    664 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 723

Qy   1881 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1940
          |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    724 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 783

Qy   1941 CTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTA 2000
          |      | ||||     |     |  |||   | |  |     || |||| |||||
Db    784 CAGTACAAACACCATTGACATCGCCACCAGCTCCGCCGACCGAGGCCCCAGCACCTACCC 843

Qy   2001 CCGCAGCACCTGCCCCCAACACCACAGTCACTG 2033
           ||  ||  || | ||     |  | |  ||||
Db    844 TCGGGGCTGCTACGCCAGGTCCAGCTGGAACTG 876


RESULT  14
US-08-298-829-25
; Sequence 25, Application US/08298829
; Patent No. 5620882
;  GENERAL INFORMATION:
;     APPLICANT:  John, Maliyakal E.
;     TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;     TITLE OF INVENTION:  PLANTS FOR ALTERED FIBER
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;       STREET:  P.O. Box 2113, First Wisconsin Plaza
;       CITY:  Madison
;       STATE:  Wisconsin
;       COUNTRY:  USA
;       ZIP:  53701
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  Microsoft Word
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/298,829
;       FILING DATE:  19-OCT-1994
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/885,970
;       FILING DATE:  18-MAY-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/617,239
;       FILING DATE:  21-NOV-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/253,243
;       FILING DATE:  04-OCT-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seay, Nicholas J.
;       REGISTRATION NUMBER:  27,386
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (608) 283-2478
;       TELEFAX:  (608) 251-5139
;   INFORMATION FOR SEQ ID NO:  25:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1985 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Gossypium barbadense
;       STRAIN:  Sea Island
;     IMMEDIATE SOURCE:
;       LIBRARY:  EMBL-SI
;       CLONE:  SIH6
US-08-298-829-25


  Query Match             3.6%;  Score 101;  DB 1;  Length 1985;
  Best Local Similarity   51.4%;  Pred. No. 8.6e-13;
  Matches  233;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;

Qy   1581 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1640
          |||||||| || |  | ||  |    || ||| | ||  |      |  | ||  | || |
Db    424 CGCCACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 483

Qy   1641 CCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1700
          ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    484 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 543

Qy   1701 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1760
            | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||
Db    544 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 603

Qy   1761 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1820
          |||   |  |  | ||| |  || | ||  | ||| | ||  |  ||||| |  |  ||
Db    604 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 663

Qy   1821 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1880
          |  || || |  ||  | |       ||| || |   | || | ||| | ||| || | |
Db    664 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 723

Qy   1881 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1940
          |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    724 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 783

Qy   1941 CTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTA 2000
          |      | ||||     |     |  |||   | |  |     || |||| |||||
Db    784 CAGTACAAACACCATTGACATCGCCACCAGCTCCGCCGACCGAGGCCCCAGCACCTACCC 843

Qy   2001 CCGCAGCACCTGCCCCCAACACCACAGTCACTG 2033
           ||  ||  || | ||     |  | |  ||||
Db    844 TCGGGGCTGCTACGCCAGGTCCAGCTGGAACTG 876
```

```
RESULT 15
US-08-217-327-3
; Sequence 3, Application US/08217327
; Patent No. 5474925
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E
;    APPLICANT:  Barton, Kenneth A
;    TITLE OF INVENTION:  Immobilized Proteins in Cotton Fiber
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O. Box 2113
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  USA
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/217,327
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/812,233
;      FILING DATE:  19-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990831
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      TISSUE TYPE:  Fiber cells
;    IMMEDIATE SOURCE:
;      CLONE:  H6
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  72..716
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  72..147
US-08-217-327-3


  Query Match             3.5%;  Score 100;  DB 1;  Length 913;
  Best Local Similarity   51.8%;  Pred. No. 1e-12;
  Matches  226;  Conservative    0;  Mismatches  210;  Indels    0;  Gaps    0;

Qy   1581 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1640
          || ||||| || |  | ||  |    || ||| | ||  |     |  | ||  | || |
Db    175 CGACACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 234

Qy   1641 CCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1700
          ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    235 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 294

Qy   1701 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1760
             | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||

Db    295 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 354

Qy   1761 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1820
          |||   |  |  | ||| |  || | || |  ||| | ||  |  ||||| |  |  ||
Db    355 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 414

Qy   1821 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1880
          |  || || |  ||  | |       ||| || |   | || | ||| | ||| || | |
Db    415 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 474

Qy   1881 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1940
          |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    475 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 534

Qy   1941 CTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTA 2000
          |      | ||||     ||    |  |||   |  |  |     || |||| |||||
Db    535 CAGTACAAACACCATTGACCTCGCCACCAGCTCCGCCGACCGAGGCCCCAGCACCTACCC 594

Qy   2001 CCGCAGCACCTGCCCC 2016
             ||  ||  || | ||
Db    595 TCGGGGCTGCTACGCC 610


Search completed: July 18, 1999, 07:54:43
Job time: 384 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 11:07:49 ; Search time 133.54 Seconds
(without alignments)
3974.226 Million cell updates/sec

Title: US-09-294-663-2
Perfect score: 2821
Sequence: 1 GAAAAGAATAACCAGCGAAC.....TAAAAAAAAAAAAAAAAAAA 2821

Scoring table: IDENTITY_NUC

Searched: 240622 seqs, 94065609 residues

Database : N_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	158.8	5.6	10596	1	Q51731	Plasmid pCisEBON f
c 2	158.8	5.6	10596	1	T40348	Plasmid pCisEBON f
c 3	158.8	5.6	9600	1	V21683	Vector plasmid pCM
c 4	158.8	5.6	799	1	V55831	Nucleotide sequenc
c 5	151.6	5.4	795	1	V55830	FLGA insert stabil
6	137.8	4.9	5163	1	V20700	Cryptosporidium pa
7	136.2	4.8	5318	1	V20701	Cryptosporidium pa
c 8	120.2	4.3	2744	1	Q98470	MiSP1-containing p
c 9	110.544	3.9	6530	1	Q51557	Loricrin gene. Con
c 10	110.4	3.9	40875	1	T80043	Insert from cosmid
11	101	3.6	1984	1	T13030	Cotton fibre-speci
12	101	3.6	1985	1	T30250	Cotton fibre clone
13	101	3.6	1985	1	T70036	Cotton H6 gene and
14	100.8	3.6	5008	1	Q27223	MSF precursor. New
15	100	3.5	913	1	T07199	Cotton fiber-speci
16	100	3.5	913	1	T13034	Cotton fibre-speci
17	100	3.5	913	1	T30253	Cotton fibre cell-
18	100	3.5	913	1	T62610	Cotton fibre speci
19	100	3.5	913	1	T70041	Cotton fibre speci
20	98.8	3.5	502	1	Q06483	Intestinal mucin c
21	97.4	3.5	1368	1	N91235	DNA sequence of pr
22	95	3.4	1107	1	V38816	DNA encoding a dur
23	93	3.3	860	1	Q06486	Intestinal mucin c
c 24	92.4	3.3	3331	1	V22683	New DNA sequence i
c 25	92.4	3.3	2214	1	V22682	New DNA sequence i
26	91.4	3.2	859	1	Q12220	SMUC-53 intestinal
27	90.8	3.2	502	1	Q12217	SMUC-40 intestinal
c 28	90.4	3.2	44377	1	T78508	Platenolide syntha
c 29	90.4	3.2	13987	1	T80415	Hybrid srmG/tylG O
c 30	90.4	3.2	44377	1	T80414	Platenolide syntha
c 31	89	3.2	2004	1	T85356	Nephila clavipes s
c 32	88.8	3.1	1505	1	Q55750	Genomic clone G11F
33	86.6	3.1	836	1	Q06484	Intestinal mucin c
34	86.6	3.1	836	1	Q12218	SMUC-41 intestinal
35	85.8	3.0	2550	1	V52497	Streptococcus pneu
c 36	85.6	3.0	2000	1	N71065	Gene encoding Plas
37	84.4	3.0	12001	1	Q76213	HSV L/ST region. H
38	83	2.9	390	1	Q21833	Randomising oligon
39	83	2.9	390	1	Q36859	PCR primer for 5'
40	82.6	2.9	1702	1	V52500	Streptococcus pneu
41	82.4	2.9	32768	1	V52204	Streptococcus pneu
42	80.4	2.9	14704	1	Q20685	PKS 741 insert con
43	80.2	2.8	4483	1	V52496	Streptococcus pneu
c 44	80	2.8	43280	1	T80413	Tylactone synthase
c 45	79.8	2.8	2338	1	V23249	Nephila clavipes s

ALIGNMENTS

RESULT 1
Q51731/c
ID Q51731 standard; DNA; 10596 BP.
AC Q51731;
DT 31-MAY-1994 (first entry)
DE Plasmid pCisEBON for subcloning huHGF variants.
KW Hepatocyte Growth Factor; HGF; variant; mutein; in vitro mutagenesis;
KW proteolysis resistant; liver; malignancy; CMV-driven;
KW Cytomegalovirus; episomal expression plasmid; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT enhancer 1. .611
FT /*tag= a
FT /note= "CMV enhancer/promoter"
FT promoter 758. .775
FT /*tag= b
FT /label= SP6_promoter
FT misc_feature 845. .849
FT /*tag= c
FT /note= "SP6 RNA start"
FT misc_feature 902. .966
FT /*tag= d
FT /function= cloning_linker
FT polya_signal 967. .1107
FT /*tag= e
FT /note= "SV40 poly A"
FT misc_feature 1108. .1531
FT /*tag= f
FT /function= SV40_origin
FT misc_feature 1580. .4189
FT /*tag= g
FT /label= EBNA-1
FT misc_feature 4190. .6374
FT /*tag= h
FT /function= oriP
FT repeat_region 4295. .4887
FT /*tag= i
FT /note= "family of repeats"
FT misc_structure 5866. .5978
FT /*tag= j
FT /note= "dyad region"
FT terminator 6375. .6457
FT /*tag= k
FT /label= HSV_TK_terminator_3'-end
FT cds 6975. .7975
FT /*tag= l
FT /phenotype= neomycin_resistance
FT /note= "Tn5 neomycin phosphotransferase gene"
FT promoter 7975. .8112
FT /*tag= m
FT /label= TK_promoter
FT misc_feature 8114. .8594
FT /*tag= n
FT /function= M13_ori
FT misc_rna 8595. .10414
FT /*tag= o
FT /label= delta_2a
PN WO9323541-A.
PD 25-NOV-1993.
PF 17-MAY-1993; U04648.
PR 18-MAY-1992; US-884811.
PR 18-MAY-1992; US-885971.
PA (GETH) GENENTECH INC.
PI Godowski PJ, Lokker NA, Mark MR;
DR WPI; 93-386573/48.
PT Hepatocyte growth factor variants - are resistant to proteolytic

```
PT   cleavage into its two-chain form, used to treat malignancies
PT   associated with HGF receptor
PS   Example 1; Fig 6; 87pp; English.
CC   Plasmid pCisEBON (a pRK5 derivative) is an episomal CMV driven
CC   expression plasmid. HuHGF variants with enhanced receptor binding
CC   activity were produced by site-directed mutagenesis. Stable
CC   populations of preferred HGF variants were obtained by transfecting
CC   human embryonic kidney 293 cells and then these were subcloned in
CC   pCisEBON. See R52940-R52949 for examples of pref. HGF variants.
SQ   Sequence   10596 BP;    2625 A;    2571 C;    3024 G;    2376 T;
```

```
  Query Match             5.6%;  Score 158.8;  DB 1;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 3.4e-20;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
Db   2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy   1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
Db   2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy   2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
Db   2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy   2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
Db   2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy   2153 GACCCCCCTCTCCC 2166
Db   2286 CCTCCTCCTGCTCC 2273
```

```
RESULT   2
T40348/c
ID   T40348 standard; DNA; 10596 BP.
AC   T40348;
DT   09-DEC-1996  (first entry)
DE   Plasmid pCisEBON for expression of hepatocyte growth factor.
KW   Human; hepatocyte growth factor; HGF; huHGF; serum; proteolytic cleavage;
KW   pro-hormone; beta subunit; alpha subunit; kringle domain; prothrombin;
KW   plasminogen; catalytic domain; serine protease; HGF variant;
KW   HGF receptor; malignancy; chronic HGF receptor activation; ss.
OS   Synthetic.
PN   US5547856-A.
PD   20-AUG-1996.
PF   18-MAY-1992; 884811.
PR   18-MAY-1992; US-885971.
PR   18-MAY-1992; US-884811.
PR   13-JUL-1993; US-087783.
PA   (GETH ) GENENTECH INC.
PI   Godowski PJ,  Lokker NA,  Mark MR;
DR   WPI; 96-392634/39.
PT   New hepatocyte growth factor variants - are resistant to in vivo
PT   proteolytic cleavage into a 2-chain form, useful as HGF antagonists
PS   Example 1; Fig 6; 39pp; English.
CC   This sequence represents the episomal CMV driven expression plasmid
CC   pCisEBON which was used in the expression of variant human hepatocyte
CC   growth factor (HGF).  HGF is isolated from human serum and is a
CC   disulphide linked heterodimer derived by proteolytic cleavage of the
CC   pro-hormone between residues 494 and 495.  This generates a molecule
CC   composed of an alpha subunit of 440 amino acids (mol. wt. 69 kD) and
CC   a beta subunit of 234 amino acids (mol. wt. 34 kD).  The alpha and beta
CC   subunits are encoded by a single open reading frame.  The alpha subunit
CC   contains four kringle domains based on their homology to kringle-like
CC   domains in other proteins, e.g. prothrombin, plasminogen.  The beta
CC   subunit shows high homology to the catalytic domain of serine proteases.
CC   However two of the three residues which form the catalytic triad of
CC   serine proteases are not conserved in HGF.  Therefore, the precise
CC   function of the beta chain remains unknown.  The invention includes HGF
CC   variants which retain HGF receptor binding activity without having the
CC   biological activity of wild-type HGF.  They can be used for the treatment
CC   of pathological conditions associated with the activation of a HGF
CC   receptor such as malignancies associated with chronic HGF receptor
CC   activation.  The pCisEBON plasmid comprising the HGF coding sequence
CC   may be used for manipulation of the HGF coding sequence and expression
CC   of the variant HGF's of the invention.
SQ   Sequence   10596 BP;    2627 A;    2571 C;    3023 G;    2375 T;
```

```
  Query Match             5.6%;  Score 158.8;  DB 1;  Length 10596;
  Best Local Similarity   53.3%;  Pred. No. 3.4e-20;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
Db   2946 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 2887

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
Db   2886 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 2827

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db   2826 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2767

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
Db   2766 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2707

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
Db   2706 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 2647

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
Db   2646 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 2587

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
Db   2586 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 2527

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
```

```
Db    2526 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 2467

Qy    1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
           || |  | ||  | ||  | ||| || |  | || ||| |  ||    |  | |   ||
Db    2466 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 2407

Qy    2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
           |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||   | |
Db    2406 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 2347

Qy    2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
           |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db    2346 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 2287

Qy    2153 GACCCCCCTCTCCC 2166
              || |||   ||
Db    2286 CCTCCTCCTGCTCC 2273
```

```
RESULT   3
V21683/c
ID   V21683 standard; DNA; 9600 BP.
AC   V21683;
DT   17-AUG-1998  (first entry)
DE   Vector plasmid pCMVKmITR-EPI.
KW   Polynucleotide delivery; plasmid pCMVKmITR-EPI; vector;
KW   gene therapy; vaccine; polycationic agent; ss.
OS   Chimeric - Epstein-Barr virus.
OS   Chimeric - Adeno-associated virus.
OS   Chimeric - Cytomegalovirus.
OS   Chimeric - Bos taurus.
FH   Key             Location/Qualifiers
FT   CDS             14. .2594
FT                   /*tag= a
FT                   /product= "EBV nuclear antigen A"
FT   misc_feature    2623. .4559
FT                   /*tag= b
FT                   /note= "EBV origin of replication"
FT   repeat_unit     4928. .5104
FT                   /*tag= c
FT                   /rpt_type= INVERTED
FT                   /note= "AAV inverted terminal repeat"
FT   repeat_unit     7189. .7355
FT                   /*tag= d
FT                   /rpt_type= INVERTED
FT                   /note= "AAV inverted terminal repeat"
FT   promoter        5112. .6734
FT                   /*tag= e
FT                   /note= "CMV immediate-early enhancer/promoter"
FT   terminator      6818. .7050
FT                   /*tag= f
FT                   /note= "bovine growth hormone polyA sequence"
PN   WO9806437-A2.
PD   19-FEB-1998.
PF   13-AUG-1997; U14465.
PR   13-AUG-1996; US-023867.
PA   (CHIR ) CHIRON CORP.
PI   Cohen F, Dubois-Stringfellow N, Dwarki V, Innis MA,
PI   Murphy JE, Tetsuo U, Zukermann R;
DR   WPI; 98-159296/14.
PT   Polycationic agents based on alpha-amino acids, able to complex
PT   with nucleic acid - to facilitate its entry into cell, condense it
PT   and protect it against serum degradation, particularly for use in
PT   gene therapy
PS   Disclosure; Page 77-80; 100pp; English.
CC   This polynucleotide comprises the DNA sequence of vector plasmid
CC   pCMVKmITR-EPI, which contains an Epstein-Barr virus (EBV) origin
CC   of replication from plasmid pCEP4, a coding region for EBV nuclear
CC   antigen A from pCEP4, a pair of inverted terminal repeats from
CC   adeno-associated virus, a cytomegalovirus enhancer/promoter, a
CC   bovine growth hormone polyA sequence, and a kanamycin resistance
CC   selectable marker.  Polynucleotides encoding polypeptides, such as
CC   erythropoietin or leptin, and ribozymes and antisense
CC   polynucleotides can be inserted into the vector.  The vector is
CC   preferred for use in novel compositions and methods for improved
CC   polynucleotide delivery into cells.  In these methods, polycationic
CC   agents are used to increase the frequency of uptake of a
CC   nucleic acid (see also V21684-86) into a cell.  The polycationic
CC   agent can condense with the nucleic acid and inhibit serum and/or
CC   nuclease degradation of the nucleic acid.  The nucleic acid can be
CC   a vector, may express a therapeutic protein or a vaccinating viral
CC   or cancer antigen, or is itself therapeutic (antisense or
CC   ribozyme).  The methods and compositions can be used in the gene
CC   therapy of many diseases.
SQ   Sequence   9600 BP;    2326 A;    2376 C;    2817 G;    2081 T;
```

```
  Query Match             5.6%;  Score 158.8;  DB 1;  Length 9600;
  Best Local Similarity   53.3%;  Pred. No. 3.3e-20;
  Matches  359;  Conservative    0;  Mismatches  312;  Indels    3;  Gaps    1;

Qy    1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
           ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db    1412 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 1353

Qy    1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
           || |  | |      | ||| |  | || || ||  ||||  | || | ||| ||| |
Db    1352 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 1293

Qy    1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
           | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db    1292 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 1233

Qy    1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
              ||||   | ||  |  ||| | |     | ||| ||||| || || |  | ||||||
Db    1232 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 1173

Qy    1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
           || || ||| ||||      ||   | ||    |  | || ||| | ||| |  || |
Db    1172 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 1113

Qy    1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
           |||||||  | ||  | ||  | ||  |  |  || |||||    |         |  ||
Db    1112 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 1053

Qy    1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
           ||| | |  |  | ||| |  || | ||  || | ||  | |||||||     || |
Db    1052 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 993

Qy    1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
           |||  ||  || || |||||||| || | ||  |  |  ||||     |  | |
Db     992 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 933

Qy    1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
           |  || |  | ||  | ||  | ||| || |  | || ||| |  ||    |  | |   ||
Db     932 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 873

Qy    2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
           |  || ||  ||||| |  || || | || || ||| ||    | ||||| ||   | |
Db     872 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 813

Qy    2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
           |  |||| || |   || |  ||    |  |  || |  | |||     || |     |
Db     812 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 753

Qy    2153 GACCCCCCTCTCCC 2166
              || |||   ||
Db     752 CCTCCTCCTGCTCC 739
```

```
RESULT   4
V55831/c
ID   V55831 standard; DNA; 799 BP.
AC   V55831;
DT   18-NOV-1998  (first entry)
DE   Nucleotide sequence of the stabilising sequence-encoding insert.
```

KW Fusion protein; stabilising polypeptide; proteolytic degradation;
KW resistance; half-life; autoimmune disease; inflammation; nitro drug;
KW IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
KW nitroreductase protein; enzyme therapy; prodrug therapy; protease;
KW cancer; pathological condition; ss.
OS Epstein-barr virus.
PN WO9822577-A1.
PD 28-MAY-1998.
PF 17-NOV-1997; IB1508.
PR 25-JUN-1997; US-048945.
PR 15-NOV-1996; US-030986.
PA (MASU/) MASUCCI M G.
PI Masucci MG;
DR WPI; 98-312463/27.
PT New fusion proteins resistant to proteolytic degradation -
PT comprising a core protein with a stabilising polypeptide comprising
PT a peptide sequence containing glycine repeats
PS Disclosure; Fig 4B; 120pp; English.
CC This is a nucleotide sequence of the stabilising sequence-encoding
CC insert. The invention provides a method for increasing the resistance
CC of a core protein to proteolytic degradation that comprises linking or
CC inserting onto or into the core protein a stabilising polypeptide of
CC formula [(Glya)X(Glyb)Y(Glyc)Z]n where Glya, Glyb, Glyc are 1-6
CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC be identical from n repeat to n repeat. Alternatively a nucleic acid
CC encoding the stabilising polypeptide can be linked onto or inserted into
CC a nucleic acid encoding a core protein. The fusion proteins of the
CC invention are more resistant to degradation by proteases and, thus, have
CC a longer half-life than the unfused core protein. The products can be
CC used for treating autoimmune diseases, cancer and inflammation. In
CC particular, the core protein may be an IkappaB regulator protein for the
CC treatment of inflammatory bowel disease, or a nitroreductase protein
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC or other pathological conditions. The fusion proteins can also be used in
CC diagnostic methods such as in vivo imaging.
SQ Sequence 799 BP; 201 A; 106 C; 479 G; 13 T;

Query Match 5.6%; Score 158.8; DB 1; Length 799;
Best Local Similarity 53.3%; Pred. No. 1.9e-20;
Matches 359; Conservative 0; Mismatches 312; Indels 3; Gaps 1;

```
Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
          ||  || || || |  | ||  | ||  |  |||||||  |  |||| ||  |  |  |
Db    741 CCTCCACCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCT 682

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
          || |  | |      | ||| |  | || || ||  ||||  | || | ||| ||| |
Db    681 GCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 622

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
           | ||   |||| | ||  | ||  |  |||||||  | ||  ||||| || || |||||
Db    621 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 562

Qy   1676 ---CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1732
             ||||  | ||  |  ||| | |     | ||| ||||| || || |  | ||||||
Db    561 CCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 502

Qy   1733 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1792
          || || ||| ||||      ||   | ||    |  | || ||| | ||| |  || |
Db    501 CCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCT 442

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1852
          |||||||  | ||  | ||  | ||  |  |  || |||||    |        |  ||
Db    441 CCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCT 382

Qy   1853 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1912
          |||  |  |   |  | |||  |  || | ||  || | ||  | ||||||||     || |
Db    381 CCTGCCCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 322

Qy   1913 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTG 1972
             |||  ||  || || |||||||| || | ||  |  |  ||||     |  | |
```

Db 321 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCC 262

Qy 1973 ACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACT 2032
|| | | || | || | ||| || | | || ||| | || | | | ||
Db 261 CCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCT 202

Qy 2033 GTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACC 2092
| || || |||| | || || | || || ||| || | ||||| || | |
Db 201 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 142

Qy 2093 ACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATC 2152
| |||| || | || | || | | || | | ||| || | |
Db 141 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCC 82

Qy 2153 GACCCCCCTCTCCC 2166
|| ||| ||
Db 81 CCTCCTCCTGCTCC 68

RESULT 5
V55830/c
ID V55830 standard; DNA; 795 BP.
AC V55830;
DT 18-NOV-1998 (first entry)
DE FLGA insert stabilising polypeptide encoding DNA.
KW Fusion protein; stabilising polypeptide; proteolytic degradation;
KW resistance; half-life; autoimmune disease; inflammation; nitro drug;
KW IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
KW nitroreductase protein; enzyme therapy; prodrug therapy; protease;
KW cancer; pathological condition; ss.
OS Epstein-barr virus.
FH Key Location/Qualifiers
FT CDS 1. .788
FT /*tag= a
FT /product= "stabilising polypeptide""
PN WO9822577-A1.
PD 28-MAY-1998.
PF 17-NOV-1997; IB1508.
PR 25-JUN-1997; US-048945.
PR 15-NOV-1996; US-030986.
PA (MASU/) MASUCCI M G.
PI Masucci MG;
DR WPI; 98-312463/27.
DR P-PSDB; W79128.
PT New fusion proteins resistant to proteolytic degradation -
PT comprising a core protein with a stabilising polypeptide comprising
PT a peptide sequence containing glycine repeats
PS Disclosure; Fig 3; 120pp; English.
CC This DNA encodes a stabilising polypeptide and is the FLGA insert of the
CC invention. The invention provides a method for increasing the resistance
CC of a core protein to proteolytic degradation that comprises linking or
CC inserting onto or into the core protein a stabilising polypeptide of
CC formula [(Glya)X(Glyb)Y(Glyc)Z]n where Glya, Glyb, Glyc are 1-6
CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC be identical from n repeat to n repeat. Alternatively a nucleic acid
CC encoding the stabilising polypeptide can be linked onto or inserted into
CC a nucleic acid encoding a core protein. The fusion proteins of the
CC invention are more resistant to degradation by proteases and, thus, have
CC a longer half-life than the unfused core protein. The products can be
CC used for treating autoimmune diseases, cancer and inflammation. In
CC particular, the core protein may be an IkappaB regulator protein for the
CC treatment of inflammatory bowel disease, or a nitroreductase protein
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC or other pathological conditions. The fusion proteins can also be used in
CC diagnostic methods such as in vivo imaging.
SQ Sequence 795 BP; 200 A; 104 C; 478 G; 13 T;

Query Match 5.4%; Score 151.6; DB 1; Length 795;
Best Local Similarity 52.0%; Pred. No. 3.6e-19;
Matches 388; Conservative 0; Mismatches 354; Indels 4; Gaps 2;

```
Qy   1403 AGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGCTGGTTGCCAAGTTTTCGAGTGC 1462
Db    777 ATTCCTGCAGCCCCGGCCTCCACTGCCTCCTCGACCCCGGCCTCCACCTCCTGCTCCTGC 718

Qy   1463 GATTCTGACAACCAGTGCACATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCA 1522
Db    717 CCCTCCTGCTCCTGCCCCTCCTCCTCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCT 658

Qy   1523 ACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACG 1582
Db    657 CCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGC---CCCTCCTGCCCCT 601

Qy   1583 CCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACC 1642
Db    600 CCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCT 541

Qy   1643 GCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTA 1702
Db    540 CCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCT 481

Qy   1703 CCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACT 1762
Db    480 GCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCT 421

Qy   1763 GTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCC 1822
Db    420 CCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCT 361

Qy   1823 CCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCC 1882
Db    360 GCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCT 301

Qy   1883 ACTACCGCAGCACCTGCCCCCAACACCACAGTCACTG-TACCACCCACTGCTGCCCCTAC 1941
Db    300 CCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCCCCCTCCTGCCCC 241

Qy   1942 TACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTAC 2001
Db    240 TCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGC 181

Qy   2002 CGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGC 2061
Db    180 CCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCC 121

Qy   2062 TCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGC 2121
Db    120 TGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCC 61

Qy   2122 ACCAGCTACCACACCTGAAGATGATG 2147
Db      60 TCCTGCTCCTGCCCCTCCCGCTCCTG 35
```

```
RESULT   6
V20700
ID   V20700 standard; DNA; 5163 BP.
AC   V20700;
DT   17-AUG-1998  (first entry)
DE   Cryptosporidium parvum GP900 antigen open reading frame.
KW   Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW   antibody; prophylaxis; treatment; inhibition; retardation;
KW   detection; diagnosis; human; ds.
OS   Cryptosporidium parvum.
FH   Key             Location/Qualifiers
FT   CDS             1. .5163
FT                   /*tag= a
FT                   /product= GP900 antigen
FT   misc_feature    524. .1270
FT                   /*tag= b
FT                   /note= "region containing NINC mutations"
PN   WO9806430-A1.
PD   19-FEB-1998.
PF   11-AUG-1997; U14104.
PR   14-AUG-1996; US-700651.
PA   (REGC ) UNIV CALIFORNIA.
PI   Gut J, Leech J, Nelson RC, Petersen C;
DR   WPI; 98-159290/14.
DR   P-PSDB; W48299.
PT   Anti-Cryptosporidium antibody - used to develop products for
PT   detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT   infections
PS   Claim 32; Pages 60-62; 89pp; English.
CC   The sequence is that encoding the GP900 antigen which may be used
CC   in the production of anti-Cryptosporidium antibodies. These can be
CC   used for the prophylaxis, treatment, inhibition or retardation of
CC   a Cryptosporidium infection in humans or in animals such as calves.
CC   They can also be used for the detection and diagnosis of related
CC   infections.
SQ   Sequence   5163 BP;    1875 A;    1137 C;    873 G;    1278 T;
```

```
  Query Match             4.9%;  Score 137.8;  DB 1;  Length 5163;
  Best Local Similarity   51.0%;  Pred. No. 1.7e-16;
  Matches  325;  Conservative    0;  Mismatches  312;  Indels    0;  Gaps    0;

Qy   1481 ACATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCT 1540
Db    584 ACAACAACAACTACTACTACTACTACTACGACAACAACAACAACAACGACAACAACAACA 643

Qy   1541 GCCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCC 1600
Db    644 ACTACTACAACTACTACCACTACTACTACGACAACAACAACAACAACAACAACAACAACA 703

Qy   1601 CCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCT 1660
Db    704 ACAACAACAACAACAACAACAACAACAACAACGACTACTACTACTACTACTACTACTACT 763

Qy   1661 CCTACTACTGCTGCCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCC 1720
Db    764 ACTACTACTACTACTACAACCACAACAACTACAACCACAACTACAACCACAACAACTACA 823

Qy   1721 ACCGCAGCCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACC 1780
Db    824 ACCACAACAACAACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACC 883

Qy   1781 GCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCA 1840
Db    884 ACAACAACCACAACCACAACCACAACTACCAAGAAACCAACAACAACAACTACAACAACA 943

Qy   1841 GAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCC 1900
Db    944 ACAACAACAACAACAACTACTACTACAACCACCACAACAACAACAACAACAACTACTACA 1003

Qy   1901 CCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCC 1960
Db   1004 ACTACCAAGAAACCAACAACTACTACTACTACTACCACAACAACAACAACTACTACTACT 1063

Qy   1961 AACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAAC 2020
Db   1064 ACCACAACAACAACAACTACTACTACTACTACAACAACAACAACAACAACAACAACAACA 1123

Qy   2021 ACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCT 2080
Db   1124 ACAACTACCACGAAACCAACAACAACAACAACAACTACTACTACTACTACAACCAAGAAA 1183

Qy   2081 GCACCCAACACCACAGCTGCCCCGGTAACTACAACCA 2117
Db   1184 CCAACAACAACTACCACTGCCACAACAACAACTACTA 1220
```

```
RESULT   7
V20701
ID   V20701 standard; DNA; 5318 BP.
AC   V20701;
DT   17-AUG-1998  (first entry)
DE   Cryptosporidium parvum GP900 antigen open reading frame and 3' region.
```

```
KW   Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW   antibody; prophylaxis; treatment; inhibition; retardation;
KW   detection; diagnosis; human; 3' region; ds.
OS   Cryptosporidium parvum.
FH   Key             Location/Qualifiers
FT   CDS             1. .5166
FT                   /*tag= a
FT                   /product= GP900 antigen
FT   misc_feature    524. .1270
FT                   /*tag= b
FT                   /note= "region containing NINC mutations"
PN   WO9806430-A1.
PD   19-FEB-1998.
PF   11-AUG-1997; U14104.
PR   14-AUG-1996; US-700651.
PA   (REGC ) UNIV CALIFORNIA.
PI   Gut J, Leech J, Nelson RC, Petersen C;
DR   WPI; 98-159290/14.
DR   P-PSDB; W48299.
PT   Anti-Cryptosporidium antibody - used to develop products for
PT   detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT   infections
PS   Claim 32; Pages 62-63; 89pp; English.
CC   The sequence is that encoding the GP900 antigen which may be used
CC   in the production of anti-Cryptosporidium antibodies. These can be
CC   used for the prophylaxis, treatment, inhibition or retardation of
CC   a Cryptosporidium infection in humans or in animals such as calves.
CC   They can also be used for the detection and diagnosis of related
CC   infections.
SQ   Sequence   5318 BP;    1944 A;    1153 C;    888 G;    1333 T;

  Query Match             4.8%;  Score 136.2;  DB 1;  Length 5318;
  Best Local Similarity   50.9%;  Pred. No. 3.3e-16;
  Matches  324;  Conservative    0;  Mismatches  313;  Indels    0;  Gaps    0;

Qy   1481 ACATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCT 1540
Db    584 ACAACAACAACTACTACTACTACTACTACGACAACAACAACAACAACGACAACAACAACA 643

Qy   1541 GCCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCC 1600
Db    644 ACTACTACAACTACTACCACTACTACTACGACAACAACAACAAAAACAACAACAACAACA 703

Qy   1601 CCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCT 1660
Db    704 ACAACAACAACAACAACAACAACAACAACGACTACTACTACTACTACTACTACTACT 763

Qy   1661 CCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCC 1720
Db    764 ACTACTACTACTACTACAACCACAACAACTACAACCACAACTACAACCACAACAACTACA 823

Qy   1721 ACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACC 1780
Db    824 ACCACAACAACAACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACC 883

Qy   1781 GCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCA 1840
Db    884 ACAACAACCACAACCACAACCACAACTACCAAGAAACCAACAACAACAACTACAACAACA 943

Qy   1841 GAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCC 1900
Db    944 ACAACAACAACAACAACTACTACTACAACCACCACAACAACAACAACAACAACTACA 1003

Qy   1901 CCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCC 1960
Db   1004 ACTACCAAGAAACCAACAACTACTACTACTACTACCACAACAACAACAACTACTACTACT 1063

Qy   1961 AACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAAC 2020
Db   1064 ACCACAACAACAACAACTACTACTACTACTACAACAACAACAACAACAACAACAACAACA 1123

Qy   2021 ACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCT 2080
Db   1124 ACAACTACCACGAAACCAACAACAACAACAACAACTACTACTACTACTACAACCAAGAAA 1183

Qy   2081 GCACCCAACACCACAGCTGCCCCGGTAACTACAACCA 2117
Db   1184 CCAACAACAACTACCACTGCCACAACAACAACTACTA 1220


RESULT   8
Q98470/c
ID   Q98470 standard; cDNA; 2744 BP.
AC   Q98470;
DT   01-MAY-1996  (first entry)
DE   MiSP1-containing plasmid pMISS1.
KW   Spider silk; repeat unit; consensus; minor ampullate silk protein;
KW   spidroin; MiSP; orb web spider; dragline; ds.
OS   Nephila clavipes.
FH   Key             Location/Qualifiers
FT   cds             183. .2678
FT                   /*tag= a
FT   misc_difference 996. .1037
FT                   /*tag= b
FT                   /note= "represented as indeterminate, a highly
FT                   compressed GC rich region which could
FT                   not be sequenced"
PN   WO9525165-A1.
PD   21-SEP-1995.
PF   14-MAR-1995; U03139.
PR   14-MAR-1994; US-209747.
PA   (UYWY-) UNIV WYOMING.
PI   Colgin M,  Lewis RV;
DR   WPI; 95-336970/43.
DR   P-PSDB; R80168.
PT   Polypeptide(s) comprising repeated unit amino acid sequences, also
PT   cDNAs - derived from minor ampullate spider silk proteins and used
PT   to form spider silk fibres
PS   Claim 12; Fig 1; 86pp; English.
CC   cDNA clone, pMISS1, encoding the orb web spider minor ampullate silk
CC   protein MiSP1 has been identified and sequenced. Repeat unit peptides
CC   of MiSP1 may form part of a larger polypeptide with an amino terminus
CC   (see R80184-85) and a carboxy terminus (see R80186-89). Repeat unit
CC   peptides make up spider silk proteins (spidroins) which in turn
CC   aggregate to form the silk fibres. Spider silk fibres have high
CC   tensile strength and significant elasticity. An isolated cDNA clone of
CC   a silk protein encoding sequence is of use to produce the protein at
CC   high yields using recombinant DNA technology.
SQ   Sequence   2744 BP;    486 A;    549 C;    1059 G;    608 T;

  Query Match             4.3%;  Score 120.2;  DB 1;  Length 2744;
  Best Local Similarity   50.1%;  Pred. No. 2.1e-13;
  Matches  333;  Conservative    0;  Mismatches  323;  Indels    9;  Gaps    1;

Qy   1482 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1541
Db   2121 CAGCGGCTCCAGCTCCTGCAGCGGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGC 2062

Qy   1542 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1601
Db   2061 CTCCGGATCCTGCACCAGCAGCTGCAGCAGCACCAGCTCCTGCTCCGGCACCGTATCCAC 2002

Qy   1602 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1661
Db   2001 CTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCAC 1942

Qy   1662 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1721
Db   1941 CAGCGGCTCCTGCACGACCACCTCTACTGTAACCCGCAGCGCCTCCAGCTCCTGCTCCAG 1882

Qy   1722 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1781
Db   1881 CAGCTGCCGCAGCACCAGCTCTTGCTCCGGCACCGTATCCACCCTGACCACCGTAGCCTC 1822
```

```
Qy   1782 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1841
Db   1821 CAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCACCTGCAGCAGCTCCAGCTC 1762

Qy   1842 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1901
Db   1761 CAGCACCAGCACCTCTACCGTAACCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCG 1702

Qy   1902 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCA---------C 1952
Db   1701 CACCAGCTCCTGCTCCGGCACCATAACCGCCTTGTCCACCATATCCAGCACCACCGGTGC 1642

Qy   1953 CTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTG 2012
Db   1641 CTGCACCTGCCCCAGCAGCGGCACCAGCTCCAGCACCAGCACCTCTACCGTAACCACCAG 1582

Qy   2013 CCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCG 2072
Db   1581 CGCCTCCGGCTCCTGTTGCAGCAGCTGCAGCAGCACCAGCTCCTGCTCCGGCACCGTATC 1522

Qy   2073 TCGCACCTGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCA 2132
Db   1521 CACCTTGACCACCGTAGCCTCCAGCTCCTGCGCCTGCAGCAGCTCCAGCTCCAGCACCAG 1462

Qy   2133 CACCT 2137
Db   1461 CACCT 1457
```

```
RESULT   9
Q51557/c
ID   Q51557 standard; DNA; 6530 BP.
AC   Q51557;
DT   17-MAY-1994  (first entry)
DE   Loricrin gene.
KW   Loricrin; gene; constitutive; inducible; vector; 5' flanking region;
KW   intron; 3' flanking region; TATA box; cap site; intron/exon boundary;
KW   DNA cassette; linker; restriction site; induction; human; psoriasis;
KW   epidermal cell; wound healing; ulcers; growth factor; antisense RNA;
KW   transforming growth factor beta; cytokine; skin; tracheal epithelia;
KW   cancer; E6; E7; papilloma virus; p53 protein; vaccine; oral epithelia;
KW   capsid protein; squamous epithelium; viral; oesophageal epithelia;
KW   vaginal epithelia; cornal epithelia; transduction; transplant; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_rna        1. .1540
FT                   /*tag=  a
FT                   /note= "5' flanking region"
FT   intron          1587. .2677
FT                   /*tag=  b
FT                   /note= "Intron and intron/exon boundary"
FT   misc_rna        4384. .6530
FT                   /*tag=  c
FT                   /note= "3' flanking region"
PN   WO9322431-A.
PD   11-NOV-1993.
PF   28-APR-1993; U03993.
PR   30-APR-1992; US-876286.
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
PI   Greenhalgh DA,  Roop DR,  Rothnagel JA;
DR   WPI; 93-368790/46.
PT   Constitutive and inducible vectors based on loricrin or keratin
PT   K6 genes - for selective expression in the epidermis, e.g. for
PT   wound healing, treating psoriasis and skin cancer, etc.
PS   Claim 23; Page 26-32; 68pp; English.
CC   This sequence represents the loricrin gene and was used in the
CC   construction of the loricrin constitutive vector of the invention.
CC   The loricrin vector has a 5' flanking region of approx. 1.5 kb, an
CC   intron of approx. 1.1 kb and a 3' flanking region of approx. 1.5 kb.
CC   The 5' flanking region of the loricrin gene comprises a TATA box,
CC   cap site, first intron and intron/exon boundary, all in the correct
CC   order and position for expression of a nucleic acid cassette.  The
CC   loricrin vector also comprises a 3' flanking region of the loricrin
CC   gene and a linker with a unique restriction site at the position of
CC   the start and stop codons, joining the two flanking sequences and
CC   also having a position for insertion of the DNA cassette (see also
CC   Q51558).  Vectors such as this can be used for; (a) in vitro induction
CC   of human epidermal cells, especially for stimulating healing of
CC   wounds, surgical incisions and ulcers, (where the vector expresses a
CC   growth factor), (b) treating psoriasis, (the vector expressing
CC   transforming growth factor beta, a soluble cytokine or antisense
CC   RNA), (c) treating skin cancer, (the vector expressing antisense RNA
CC   of the E6 or E7 gene of human papilloma virus or normal p53 protein),
CC   (d) for vaccination (the vector expresses a viral capsid protein;
CC   especially of human papilloma virus), or (e) treating cancer of the
CC   squamous epithelium (of epidermal, oral, oesophageal, vaginal,
CC   tracheal or cornal epithelia, the vector expresses antisense RNA).
CC   Alternatively human epithelial cells are transduced ex vivo, then
CC   transplanted.
SQ   Sequence   6530 BP;    1635 A;    1507 C;    1696 G;    1692 T;
```

```
  Query Match             3.9%;  Score 110.544;  DB 1;  Length 6530;
  Best Local Similarity   48.4%;  Pred. No. 1.4e-11;
  Matches  307;  Conservative    0;  Mismatches  327;  Indels    0;  Gaps    0;

Qy   1500 CAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCT 1559
Db   3490 CCGCCGCCGCCACCGGAGGAACCACCTCCGCAGCTCGTGCCGCCGCCACCGGAATAGCCG 3431

Qy   1560 CCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCG 1619
Db   3430 CCGCCGCCGCCGCAGCTGGATCCACCGCCACCTGAGTAGCCGCCGCCACAGCTAGAGCCG 3371

Qy   1620 CAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTG 1679
Db   3370 CCGCTGCTTCCACCGCCGCCGGAATAGCCGCCTCCGCAGCTGGAGCCACCGCCGCCACCA 3311

Qy   1680 AATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTG 1739
Db   3310 GAGTACTTGCCGCCTCCGGAACCGCCGCCGCAGCCACCGGAGCTGCCGCCTCCAGAGGAG 3251

Qy   1740 CTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCC 1799
Db   3250 CCTCCGCAGTAGGAGCCGCCGCCTCCTGATCCGCCTCCGCAGCTGGAGCCGCCGCCGCCG 3191

Qy   1800 CCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCA 1859
Db   3190 CCCCCGGAGGAGCCGCCGCCGCAGCTAGAGCCGCCGCCGCCACCACCTCCGGAGTACTTG 3131

Qy   1860 CATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTG 1919
Db   3130 ACGCCCCCACCAGAGCCGCCACCGCCTCCACAGCTGGAGCCACCCCCGGAGTAGCTGCCG 3071

Qy   1920 TACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCAC 1979
Db   3070 CCGCCGCAGCTGGAGCCGCCGCCGCCTCCGGAGTAGCCACCTCCGCAGCTGGAGCCGCCG 3011

Qy   1980 CACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCAC 2039
Db   3010 CCGCCCCCAGAGTAGCCACCGCCGCAGCTAGAGCCACCACCGCCTCCGGAGTAGCCGCCG 2951

Qy   2040 CCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTG 2099
Db   2950 CCGCAGCTAGAGCCACCGCCGCCTCCGGAGTACTTGACGCCCCCACCGGAGCCGCCTCCA 2891

Qy   2100 CCCCGGTAACTACAACCAGCGCACCAGCTACCAC 2133
Db   2890 CCGCCGCAGCTGGAACCACCTCCATAGGAACCAC 2857
```

```
RESULT  10
T80043/c
ID   T80043 standard; DNA; 40875 BP.
AC   T80043;
DT   03-NOV-1997  (first entry)
```

```
DE   Insert from cosmid 109.
KW   Pectinase; yeast; polygalacturonase; phytopathogen; vegetable spoilage;
KW   fruit spoilage; food processing; fruit juice; cosmid; ss.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   cds             312. .1943
FT                   /*tag= a
FT                   /note= "indicated in the specification as an open reading
FT                   frame, encoding a protein with homology to
FT                   S. cerevisiae allantoate permease (DAL5)"
FT   cds             3622. .7107
FT                   /*tag= b
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             3965. .5188
FT                   /*tag= c
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             3461. .4546
FT                   /*tag= d
FT                   /note= "indicated in the specification as an open reading
FT                   frame, encoding a protein with homologies to
FT                   pectinases"
FT   cds             5151. .6194
FT                   /*tag= e
FT                   /note= "indicated in the specification as an open reading
FT                   frame, with homology to S. pombe thiamine
FT                   repressed gene"
FT   cds             5486. .5845
FT                   /*tag= f
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             6386. .7414
FT                   /*tag= g
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             6430. .7470
FT                   /*tag= h
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             8051. .8917
FT                   /*tag= i
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             9656. .10552
FT                   /*tag= j
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             9996. .10364
FT                   /*tag= k
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             10127. .10936
FT                   /*tag= l
FT                   /note= "indicated in the specification as an open reading
FT                   frame, encoding a protein with homology to the
FT                   S. cerevisiae mitochondrial genome maintenance
FT                   protein (MGM101)"
FT   cds             11161. .11523
FT                   /*tag= m
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             13441. .14517
FT                   /*tag= n
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             14058. .14360
FT                   /*tag= o
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             14989. .16119
FT                   /*tag= p
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             16405. .17619
FT                   /*tag= q
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             16491. .17336
FT                   /*tag= r
FT                   /note= "indicated in the specification as an open reading
FT                   frame, encoding a protein with homology to the
FT                   S. cerevisiae ribosomal protein S7A (SCRPS7A)"
FT   cds             18429. .18797
FT                   /*tag= s
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             18749. .21037
FT                   /*tag= t
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             21763. .22488
FT                   /*tag= u
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             22540. .23262
FT                   /*tag= v
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             23667. .28607
FT                   /*tag= w
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
PN   WO9704106-A2.
PD   06-FEB-1997.
PF   15-JUL-1996; E03091.
PR   22-DEC-1995; EP-870137.
PR   14-JUL-1995; EP-870089.
PA   (INNO-) INNOGENETICS NV.
PI   Scarcez T,  Van Broekhoven A;
DR   WPI; 97-132645/12.
PT   New nucleic acid encoding Saccharomyces pectinase - useful in food
PT   processing, also its regulatory sequences are useful for expression
PT   of heterologous protein
PS   Example 1; Fig 2; 65pp; English.
CC   This sequence represents the insert from cosmid 109, containing
CC   Saccharomyces cerevisiae strain FY1679 DNA. This sequence contained the
CC   pectinase gene of the invention (see T80042). Pectinases are synthesised
CC   by plants and a variety of microorganisms, the best characterised
CC   pectinase being polygalacturonase. Pectinases play roles in the invasion
CC   of plant tissues by phytopathogens, the spoilage of fruits and
CC   vegetables, and also in their food processing and plant biological
CC   applications. Vectors containing the pectinase coding sequence are used
CC   for production of homologous or heterologous proteins in transformed host
CC   cells. The pectinase sequence may include regulatory and/or coding
CC   sequences, including the sequence encoding the pectinase signal peptide
CC   (which can be used in any recombinant system). The pectinase is useful in
CC   food processing (e.g. preparation of fruit juices), typical heterologous
CC   proteins that can be made are (poly)peptides of bacterial, animal or
CC   viral origin. Also, probes derived from the pectinase coding sequence can
CC   be used to detect the coding sequence and primers for amplification are
CC   used to isolate related genes.
SQ   Sequence   40875 BP;    12953 A;    7579 C;    8292 G;    12051 T;


  Query Match             3.9%;  Score 110.4;  DB 1;  Length 40875;
  Best Local Similarity   48.3%;  Pred. No. 2.3e-11;
  Matches  309;  Conservative    0;  Mismatches  331;  Indels    0;  Gaps    0;

Qy   1491 CTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTG 1550
           || | || ||  |  ||   ||  || |  |  |  |  |  | ||  || |  ||||
Db  24579 CTACTCCTACCACCACTATCACTTCTACTACTTCTACCACATCGACTACTCCTACAACCT 24520

Qy   1551 CCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCAC 1610
          |  |  | || |||       || || | ||  || |  | ||| |  |  |     |
Db  24519 CTACTACTTCTACCACTCCTACAACTTCTACCACATCGACTACTCCTACAACCTCTACTA 24460
```

```
Qy   1611 CTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTG 1670
          || | ||| |   | | ||| |  || |  |  || | || ||  |  || ||||| ||
Db  24459 CTTCTACCACTCCTACAACTTCTACCACATCGACTACTCCTACAACCTCTACTACTTCTA 24400

Qy   1671 CTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCC 1730
          |  | |||  | |  | |||||    ||    ||  | ||  || || |  |  |  | | |
Db  24399 CCACTCCTACAACTTCTACCACATCGACTACTCCTACAACCTCTACTACTTCTACCACTC 24340

Qy   1731 CTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCA 1790
          |||| ||| ||    ||   | | |    ||     |  |  | |||||  || |  | |
Db  24339 CTACAACTTCTACAACTTCTACCACTCCTACAACTTCTACCACTCCTACAACTTCTACAA 24280

Qy   1791 CCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAA 1850
          || |  || |  |     |  |||||  | |  ||||| || ||    ||   |  | |
Db  24279 CCTCCACCACTTCTCAAACTTCCACCAAATCGACTACTCCTACTACTTCAAGCACATCGA 24220

Qy   1851 CTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCA 1910
          |||||   ||| |  |  | ||| || |  |  ||||  |  |  | || || |  ||
Db  24219 CTACTCCTACAACTTCTACCACTCCTACAACTTCTACAACCTCTACCGCTCCTACAACTT 24160

Qy   1911 CAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAG 1970
          |    |||    || || || |  |  |||| |   |  |  || |  | | | | |||
Db  24159 CTACAACTTCTACAACCTCTACCACTTCTACCATTTCTACCGCTCCCACGACCTCTACAA 24100

Qy   1971 TGACTGCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCA 2030
             |     |  ||  |  |  | | | ||| |  ||| |     ||||| |  |
Db  24099 CTTCAAGCACTTTCAGTACTTCTAGTGCCTCTGCATCCTCTGTCATTTCTACCACTGCTA 24040

Qy   2031 CTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACA 2090
          |     |  |||||  ||||     | |  | ||  | ||     |  || |  |  ||
Db  24039 CAACCTCTACCACTTTTGCCAGTCTTACTACACCTGCCACATCTACGGCTTCTACTGACC 23980

Qy   2091 CCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTAC 2130
            ||| |  |    ||  | ||||| | ||||    ||||
Db  23979 ATACAACCTCTAGTGTCTCAACAACTAACGCATTCACTAC 23940
```

RESULT 11
T13030
ID T13030 standard; DNA; 1984 BP.
AC T13030;
DT 27-MAY-1996 (first entry)
DE Cotton fibre-specific H6 gene and promoter.
KW Cotton; fibre; promoter; transgenic plant; crop improvement; ds.
OS Gossypium barbadense strain Sea Island.

FH	Key	Location/Qualifiers
FT	promoter	1. .250
FT		/*tag= a
FT	5'utr	251. .320
FT		/*tag= b
FT	cds	321. .1547
FT		/*tag= c
FT		/*note= "the CDS includes a 583 bp intron
FT		(location unspecified)"
FT	3'utr	1548. .1984
FT		/*tag= d

PN US5495070-A.
PD 27-FEB-1996.
PF 04-OCT-1988; 253243.
PR 04-OCT-1988; US-253243.
PR 21-NOV-1990; US-617239.
PR 18-MAY-1992; US-885970.
PA (CETU) AGRACETUS INC.
PI John M;
DR WPI; 96-139095/14.
PT New isolated fibre-specific promoters - used for introducing
PT altered fibre-specific characteristics into plants, partic. cotton.
PS Claim 1; Column 59-62; 48pp; English.
CC The cotton var. Sea Island H6 gene and promoter (T13030) were
CC identified in genomic clone EMBL-S1-H6-4 isolated by screening a
CC genomic library prepd. in lambda EMBL with cotton fibre-specific
CC cDNA clone H6 (see T13034). H6 and other (see T13025-29, T13031-32
CC and T13052-53) fibre-specific promoters can be utilised in the
CC construction of vectors used for the introduction of altered fibre-
CC specific characteristics into plants, partic. cotton. They can be
CC used to modulate the synthesis of fibre proteins or to introduce
CC non-fibre proteins into fibre in a tissue-specific manner.
SQ Sequence 1984 BP; 565 A; 470 C; 281 G; 668 T;

Query Match 3.6%; Score 101; DB 1; Length 1984;
Best Local Similarity 51.4%; Pred. No. 5.5e-10;
Matches 233; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

```
Qy   1581 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1640
          |||||||| || |  | ||  |    || ||| | ||  |     |  | ||  | || |
Db    424 CGCCACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 483

Qy   1641 CCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1700
          ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    484 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 543

Qy   1701 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1760
             | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||
Db    544 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 603

Qy   1761 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1820
          |||   |  |  | ||| |  || | ||  | ||| | ||  |  ||||| |  |  || |
Db    604 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 663

Qy   1821 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1880
          |  || || |  ||  | |       ||| || |   | || | ||| | ||| || | |
Db    664 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 723

Qy   1881 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1940
          |  ||||| || |||| || ||     | |     | |   ||  | | |||  | ||
Db    724 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 783

Qy   1941 CTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTA 2000
          |      | ||||     |     |  |||  .|  |  |     || |||| ||||
Db    784 CAGTACAAACACCATTGACATCGCCACCAGCTCCGCCGACCGAGGCCCCAGCACCTACCC 843

Qy   2001 CCGCAGCACCTGCCCCCAACACCACAGTCACTG 2033
            ||  ||  || | ||     |  | |  ||||
Db    844 TCGGGGCTGCTACGCCAGGTCCAGCTGGAACTG 876
```

RESULT 12
T30250
ID T30250 standard; DNA; 1985 BP.
AC T30250;
DT 16-DEC-1996 (first entry)
DE Cotton fibre clone SKSIE6-H6-RI-derived promoter contg. sequence.
KW Cotton fibre; promoter; differential screening; leaf; ovule; root;
KW flower; PCR; polymerase chain reaction; homology; transgenic plant; ds.
OS Gossypium barbadense.

FH	Key	Location/Qualifiers
FT	promoter	1. .250
FT		/*tag= a
FT		/note= "contains promoter sequence (claimed)"
FT	misc_feature	251. .321
FT		/*tag= b
FT		/note= "5' non-coding sequence"
FT	cds	322. .1547
FT		/*tag= c
FT		/note= "contains a 583 bp intron"
FT	misc_feature	1548. .1984
FT		/*tag= d
FT		/note= "3' non-coding sequence"

PN US5521078-A.
PD 28-MAY-1996.
PF 04-OCT-1988; 253243.
PR 04-OCT-1988; US-253243.
PR 21-NOV-1990; US-617239.
PR 18-MAY-1992; US-885970.

```
PR   19-OCT-1994; US-298687.
PA   (CETU ) AGRACETUS INC.
PI   John M;
DR   WPI; 96-267794/27.
PT   Isolation of fibre-specific cotton promoter sequences - using
PT   selected DNA probes to screen genomic DNA fragments, for production
PT   of cotton fibres with improved characteristics
PS   Claim 1; Column 59-62; 46pp; English.
CC   Cotton fibre cell-specific promoter sequences were isolated by
CC   differential screening of a cotton plant cDNA library.  Of 4788 clones
CC   from a 10 day cell library screened with leaf cDNAs, 800 clones not
CC   present in the leaf were isolated. These were screened with cDNAs from
CC   ovule, root and flower mRNAs and resulted in 79 clones isolated.  PCR
CC   analysis was then used to remove cross-hybridising clones.  This resulted
CC   in the isolation of 18 cDNA clones specifically expressed in cotton fibre
CC   cells (T30242-4 and T30253-67). These cDNAs were then used to screen for
CC   homologous genomic sequences (T30245-53 and T30268) in order to obtain
CC   the corresp. promoter sequences.
CC   This genomic clone is derived from the 1.9 kb SalI-EcoRI fragment of the
CC   sub-clone designated SKSIE6-H6-RI.  This forms part of the genomic clone
CC   designated as EMBL-SI-H6-4 which contains a 13 kb MboI insert.
CC   The promoters isolated from the fibre cell-specific clones can be used to
CC   generate transgenic cotton plants and lines producing fibres having
CC   altered quantity and quality.
SQ   Sequence   1985 BP;    565 A;    470 C;    282 G;    668 T;
```

```
  Query Match             3.6%;  Score 101;  DB 1;  Length 1985;
  Best Local Similarity   51.4%;  Pred. No. 5.5e-10;
  Matches  233;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;

Qy   1581 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1640
           |||||||| || |  | ||  |    || ||| | ||  |      |  | ||  | || |
Db    424 CGCCACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 483

Qy   1641 CCGCTGCCCCCCACCGCAGCTCCTACTACTGCTGCCCCCTGAATCCCCAACCACTGTCACAG 1700
           ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    484 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 543

Qy   1701 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1760
             | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||
Db    544 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 603

Qy   1761 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1820
           |||   |  |  | ||| |  || | ||  | ||| | ||  |  ||||| |  |  ||
Db    604 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 663

Qy   1821 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1880
           |  || || |  ||  | |        ||| || |  | || | ||| | ||| || | |
Db    664 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 723

Qy   1881 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1940
           |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    724 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 783

Qy   1941 CTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTA 2000
           |      | ||||     |     |  |||   |  |  |     || |||| |||||
Db    784 CAGTACAAACACCATTGACATCGCCACCAGCTCCGCCGACCGAGGCCCCAGCACCTACCC 843

Qy   2001 CCGCAGCACCTGCCCCCAACACCACAGTCACTG 2033
            ||  ||  || | ||     |  | |  ||||
Db    844 TCGGGGCTGCTACGCCAGGTCCAGCTGGAACTG 876
```

```
RESULT  13
T70036
ID   T70036 standard; DNA; 1985 BP.
AC   T70036;
DT   20-AUG-1997  (first entry)
DE   Cotton H6 gene and fibre-specific promoter from clone SIH6.
KW   cotton; E6; fibre; promoter; transgenic plant; truncated;
KW   heterologous gene expression; ds.
OS   Gossypium barbadense strain Sea Island.
```

```
PN   US5620882-A.
PD   15-APR-1997.
PF   04-OCT-1988; 253243.
PR   04-OCT-1988; US-253243.
PR   21-NOV-1990; US-617239.
PR   18-MAY-1992; US-885970.
PR   19-OCT-1994; US-298829.
PA   (CETU ) AGRACETUS INC.
PI   John M;
DR   WPI; 97-235185/21.
PT   DNA constructs contg. truncated promoter sequence  - for
PT   fibre-specific gene expression in cotton plants
PS   Claim 1; Column 61-64; 48pp; English.
CC   T70031-38 are genomic DNA clones containing cotton fibre-specific
CC   promoters. Claimed DNA constructs comprise a truncated promoter sequence
CC   (from one of T70031-38) that promotes preferential gene expression in
CC   plant fibre cells, a protein coding sequence not naturally associated
CC   with the promoter sequence and a 3' termination sequence. The DNA
CC   constructs are useful for expressing foreign genes in fibre-producing
CC   plants, esp. to produce transgenic cotton plants with varied cotton fibre
CC   characteristics and quality. The present sequence comprises a fragment of
CC   pSIH6-4R1 (an EcoRI fragment isolated from pSKSIH6-4 which hybridises to
CC   H6 cDNA).
SQ   Sequence   1985 BP;    565 A;    470 C;    282 G;    668 T;
```

```
  Query Match             3.6%;  Score 101;  DB 1;  Length 1985;
  Best Local Similarity   51.4%;  Pred. No. 5.5e-10;
  Matches  233;  Conservative    0;  Mismatches  220;  Indels    0;  Gaps    0;

Qy   1581 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1640
           |||||||| || |  | ||  |    || ||| | ||  |      |  | ||  | || |
Db    424 CGCCACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 483

Qy   1641 CCGCTGCCCCCCACCGCAGCTCCTACTACTGCTGCCCCCTGAATCCCCAACCACTGTCACAG 1700
           ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    484 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 543

Qy   1701 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1760
             | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||
Db    544 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 603

Qy   1761 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1820
           |||   |  |  | ||| |  || | ||  | ||| | ||  |  ||||| |  |  ||
Db    604 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 663

Qy   1821 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1880
           |  || || |  ||  | |        ||| || |  | || | ||| | ||| || | |
Db    664 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 723

Qy   1881 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1940
           |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    724 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 783

Qy   1941 CTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTA 2000
           |      | ||||     |     |  |||   |  |  |     || |||| |||||
Db    784 CAGTACAAACACCATTGACATCGCCACCAGCTCCGCCGACCGAGGCCCCAGCACCTACCC 843

Qy   2001 CCGCAGCACCTGCCCCCAACACCACAGTCACTG 2033
            ||  ||  || | ||     |  | |  ||||
Db    844 TCGGGGCTGCTACGCCAGGTCCAGCTGGAACTG 876
```

```
RESULT  14
Q27223
ID   Q27223 standard; DNA; 5008 BP.
AC   Q27223;
DT   02-FEB-1993  (first entry)
DE   MSF precursor.
KW   Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
KW   stability; proteolytic cleavage; adhesion; alternative splicing; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
```

```
FT   cds             1. .4215
FT                   /*tag= a
FT   exon            1. .76
FT                   /*tag= b
FT                   /number= I
FT   exon            77. .199
FT                   /*tag= c
FT                   /number= II
FT   exon            200. .319
FT                   /*tag= d
FT                   /number= III
FT   exon            320. .469
FT                   /*tag= e
FT                   /number= IV
FT   exon            470. .598
FT                   /*tag= f
FT                   /number= V
FT   exon            599. .3421
FT                   /*tag= g
FT                   /number= VI
FT   exon            3422. .3499
FT                   /*tag= h
FT                   /number= VII
FT   exon            3500. .3636
FT                   /*tag= i
FT                   /number= VIII
FT   exon            3637. .3796
FT                   /*tag= j
FT                   /number= IX
FT   exon            3797. .3991
FT                   /*tag= k
FT                   /number= X
FT   exon            3992. .4117
FT                   /*tag= l
FT                   /number= XI
FT   exon            4118. .4212
FT                   /*tag= m
FT                   /number= XII
PN   WO9213075-A.
PD   06-AUG-1992.
PF   17-JAN-1992; U00433.
PR   18-JAN-1991; US-643502.
PR   10-SEP-1991; US-757022.
PA   (GEMY ) GENETICS INST INC.
PI   Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
DR   WPI; 92-284660/34.
DR   P-PSDB; R26049.
PT   New human mega-karyocyte stimulating factors - for treating
PT   immune deficiencies, cancer, exposure to radiation or drugs,
PT   bacterial and viral infections, etc.
PS   Claim 1, 2 and 3; Fig 1; 87pp; English.
CC   The sequence covered by exons II, III and IV encodes megakaryocyte
CC   stimulating factor (MSF). This sequence is modified by the addition
CC   of an N-terminal sequence encoding a secretory leader, an initiating
CC   methionine proceeding exon II and a terminating codon following exon
CC   IV. The cDNA sequence given contains sequences derived from human
CC   megakaryocyte colony stimulating factor (meg-CSF). Exon I contains
CC   the initiating methionine, and encodes a classical mammalian protein
CC   secretion signal sequence. The sequence encoding the original meg-CSF
CC   includes exons II-IV and is thought to terminate in the region between
CC   amino acid residues 134 - 147. The primary transcript of this gene
CC   may be cleaved in different ways to yield a family of mRNA's each
CC   encoding a different MSF protein. Exons V and VI are thought to be
CC   related to the activity of the factor and are also implicated in the
CC   stability, folding and processing of the molecule. These exons are
CC   also thought to play a role in the observed synergy of MSF with other
CC   cytokines. Exons V - XII are believed to be implicated in the
CC   processing or folding of the appropriate structure of the resulting
CC   factor, ie. one or more of these exons may contain sequences which
CC   direct proteolytic cleavage, adhesion, organisation of the cellular
CC   matrix or extracellular matrix processing. Both naturally occuring
CC   and non-naturally occuring MSF's may be characterised by various
CC   combinations of alternatively spliced exons from this sequence, with
CC   the exons spliced together in differing orders to form different
CC   members of the MSF family.
SQ   Sequence   5008 BP;    1701 A;    1506 C;    797 G;    1004 T;


  Query Match             3.6%;  Score 100.8;  DB 1;  Length 5008;
  Best Local Similarity   46.8%;  Pred. No. 7.3e-10;
  Matches  318;  Conservative    0;  Mismatches  362;  Indels    0;  Gaps    0;

Qy   1496 CCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCA 1555
Db   1096 CCCACACCTACCACCATCAAGTCTGCACCCACCACCCCCAAGGAGCCTGCACCCACCACC 1155

Qy   1556 CCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACA 1615
Db   1156 ACCAAGTCTGCACCCACCACTCCCAAGGAGCCTGCACCCACCACCACCAAGGAGCCTGCA 1215

Qy   1616 ACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCC 1675
Db   1216 CCCACCACTCCCAAGGAGCCTGCACCCACCACCACCAAGGAGCCTGCACCCACCACCACC 1275

Qy   1676 CCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACT 1735
Db   1276 AAGTCTGCACCCACCACTCCCAAGGAGCCTGCACCCACCACCCCCAAGAAGCCTGCCCCA 1335

Qy   1736 ACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCT 1795
Db   1336 ACTACCCCCAAGGAGCCTGCACCCACCACTCCCAAGGAGCCTACACCCACCACTCCCAAG 1395

Qy   1796 GCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACT 1855
Db   1396 GAGCCTGCACCCACCACCAAGGAGCCTGCACCCACCACTCCCAAAGAGCCTGCACCCACT 1455

Qy   1856 GTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTC 1915
Db   1456 GCCCCCAAGAAGCCTGCCCCAACTACCCCCAAGGAGCCTGCACCCACCACTCCCAAGGAG 1515

Qy   1916 ACTGTACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACT 1975
Db   1516 CCTGCACCCACCACCACCAAGGAGCCTTCACCCACCACTCCCAAGGAGCCTGCACCCACC 1575

Qy   1976 GCACCACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTA 2035
Db   1576 ACCACCAAGTCTGCACCCACCACTACCAAGGAGCCTGCACCCACCACTACCAAGTCTGCA 1635

Qy   2036 CCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACA 2095
Db   1636 CCCACCACTCCCAAGGAGCCTTCACCCACCACCACCAAGGAGCCTGCACCCACCACTCCC 1695

Qy   2096 GCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGAC 2155
Db   1696 AAGGAGCCTGCACCCACCACCCCCAAGAAGCCTGCCCCAACTACCCCCAAGGAGCCTGCA 1755

Qy   2156 CCCCCTCTCCCCAACGACCC 2175
Db   1756 CCCACCACTCCCAAGGAACC 1775


RESULT  15
T07199
ID   T07199 standard; DNA; 913 BP.
AC   T07199;
DT   10-MAY-1996  (first entry)
DE   Cotton fiber-specific H6 protein gene.
KW   H6 gene; promoter; cotton; transgenic plant; fiber; immobilisation;
KW   enzyme; pesticide degradation; parathion hydrolase; extensin; ds.
OS   Gossypium hirsutum strain Coker 312.
FH   Key             Location/Qualifiers
FT   cds             72. .716
FT                   /*tag= a
FT   signal_peptide  72. .147
FT                   /*tag= b
FT   mat_peptide     148. .713
```

```
FT                   /*tag=  c
PN   US5474925-A.
PD   12-DEC-1995.
PF   19-DEC-1991; 812233.
PR   19-DEC-1991; US-812233.
PR   23-MAR-1994; US-217327.
PA   (CETU ) AGRACETUS INC.
PI   Barton KA,  Maliyakal J;
DR   WPI; 96-039513/04.
DR   P-PSDB; R86913.
PT   Prepn. of immobilised enzymes in cotton fibre - esp.
PT   pesticide-degrading enzymes e.g. parathion hydrolase
PS   Disclosure; Column 31-34; 37pp; English.
CC   The cotton H6 gene (T07198) encodes an extensin-like protein
CC   (R86913) that is an integral part of the secondary cell wall
CC   of cotton.  The portions of the gene encoding the N-terminal
CC   or transmembrane regions of the protein may be utilised in the
CC   construction of gene fusions used to anchor (i.e. immobilise)
CC   foreign proteins in the fibers of transgenic cotton plants.
CC   Fusions of H6 and a foreign protein should be localised within
CC   the secondary walls of the fiber.  The immobilised proteins,
CC   e.g. enzymes or antibodies, have industrial, medical and
CC   environmental applns.
SQ   Sequence   913 BP;    196 A;    313 C;    155 G;    249 T;


  Query Match             3.5%;  Score 100;  DB 1;  Length 913;
  Best Local Similarity   51.8%;  Pred. No. 6.9e-10;
  Matches  226;  Conservative    0;  Mismatches  210;  Indels    0;  Gaps    0;

Qy   1581 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1640
            || ||||| || |  | ||  |    || ||| | ||  |     |  | ||  | || |
Db    175 CGACACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 234

Qy   1641 CCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1700
            ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    235 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 294

Qy   1701 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1760
               | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||
Db    295 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 354

Qy   1761 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1820
            |||   |  |  | ||| |  || | ||  | ||| | ||  |  ||||| |  |  ||
Db    355 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 414

Qy   1821 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1880
            |  || || |  ||  | |       ||| || |   | || | ||| | ||| || | |
Db    415 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 474

Qy   1881 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1940
            |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    475 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 534

Qy   1941 CTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACCACCCACCGCAGCCCCTACTA 2000
            |      | ||||     ||    |  |||   |  |  |     || |||| |||||
Db    535 CAGTACAAACACCATTGACCTCGCCACCAGCTCCGCCGACCGAGGCCCCAGCACCTACCC 594

Qy   2001 CCGCAGCACCTGCCCC 2016
            ||  ||  || | ||
Db    595 TCGGGGCTGCTACGCC 610


Search completed: July 18, 1999, 11:08:38
Job time: 2188 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 11:07:32 ; Search time 1043.53 Seconds
(without alignments)
3588.701 Million cell updates/sec

Title: US-09-294-663-1
Perfect score: 2455
Sequence: 1 GTAACGTTAAGTGAAAAGAA.....AAAAAAAAAAAAAAAAAAAA 2455

Scoring table: IDENTITY_NUC

Searched: 2002476 seqs, 762712212 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: gb_est1:*
11: gb_est2:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: gb_est6:*
16: gb_est7:*
17: gb_est8:*
18: gb_est9:*
19: gb_est10:*
20: gb_est11:*
21: gb_est12:*
22: gb_est13:*
23: gb_est22:*
24: gb_est14:*
25: gb_est15:*
26: gb_est16:*
27: gb_est17:*
28: gb_est18:*
29: gb_est19:*
30: gb_est20:*
31: gb_est21:*
32: em_est10:*
33: em_est11:*
34: em_est12:*
35: em_est13:*
36: em_est14:*
37: em_est15:*
38: em_est16:*
39: em_est17:*
40: em_est18:*
41: em_est19:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	110	4.5	538	30	AI114969	AI114969 ui41b06.y
	2	107.8	4.4	814	30	AI131654	AI131654 ui41b06.x
c	3	96.8	3.9	1075	14	W29419	W29419 mb98h11.r1
c	4	94	3.8	967	14	W34662	W34662 mc32c11.r1
	5	90.4	3.7	616	31	AU034820	AU034820 Dictyoste
c	6	90.2	3.7	452	22	AA589477	AA589477 v147h12.s
c	7	90.2	3.7	514	26	AA798859	AA798859 vv94a07.r
	8	90	3.7	672	31	AU034104	AU034104 Dictyoste
c	9	89.6	3.6	482	24	AA664125	AA664125 ac06c04.s
	10	89.4	3.6	690	31	AU033402	AU033402 Dictyoste
	11	89.4	3.6	566	31	AU033450	AU033450 Dictyoste
	12	89.4	3.6	574	31	AU033594	AU033594 Dictyoste
	13	89.4	3.6	664	31	AU033617	AU033617 Dictyoste
	14	89.4	3.6	603	31	AU033659	AU033659 Dictyoste
	15	89.4	3.6	586	31	AU033708	AU033708 Dictyoste
	16	89.4	3.6	578	31	AU033712	AU033712 Dictyoste
	17	89.4	3.6	672	31	AU033776	AU033776 Dictyoste
	18	89.4	3.6	605	31	AU033778	AU033778 Dictyoste
	19	89.4	3.6	638	31	AU033787	AU033787 Dictyoste
	20	89.4	3.6	622	31	AU034132	AU034132 Dictyoste
	21	89.4	3.6	565	31	AU034176	AU034176 Dictyoste
	22	89.4	3.6	773	31	AU034671	AU034671 Dictyoste
	23	89.4	3.6	630	31	AU034708	AU034708 Dictyoste
	24	89.4	3.6	575	31	AU034754	AU034754 Dictyoste
	25	89.4	3.6	686	31	AU034976	AU034976 Dictyoste
c	26	88.8	3.6	1222	13	W42209	W42209 mc69f11.r1
c	27	88.6	3.6	759	13	W28277	W28277 44g8 Human
	28	87.8	3.6	459	23	AI285778	AI285778 qu56b02.x
	29	85.8	3.5	586	31	AU033861	AU033861 Dictyoste
c	30	85.6	3.5	355	26	AA798861	AA798861 vv94a10.r
c	31	85	3.5	408	21	AA484400	AA484400 nf10e04.s
c	32	85	3.5	422	25	AA673536	AA673536 vn36g11.r
	33	84.4	3.4	555	31	AU033469	AU033469 Dictyoste
	34	84.4	3.4	521	31	AU033607	AU033607 Dictyoste
	35	84.2	3.4	558	31	AU033502	AU033502 Dictyoste
c	36	83.8	3.4	1461	14	W15723	W15723 mb53c01.r1
c	37	83.6	3.4	1257	13	W98479	W98479 mg20c12.r1
c	38	83.6	3.4	1311	14	W15735	W15735 mb53g03.r1
c	39	82.8	3.4	1161	13	W41959	W41959 mc68b10.r1
	40	82.6	3.4	498	31	AI253548	AI253548 aq30g07.x
	41	82.4	3.4	558	31	AU034078	AU034078 Dictyoste
	42	82	3.3	714	20	AA519116	AA519116 TgESTzz33
	43	81.8	3.3	636	10	T55873	T55873 yb82h02.s1
	44	80.6	3.3	484	17	AA314486	AA314486 EST186534
c	45	80.4	3.3	1199	13	W42210	W42210 mc69f12.r1

ALIGNMENTS

RESULT 1
AI114969/c
LOCUS AI114969 538 bp mRNA EST 02-SEP-1998
DEFINITION ui41b06.y1 Sugano mouse embryo mewa Mus musculus cDNA clone 1884947 5' similar to TR:O14633 O14633 SKIN-SPECIFIC PROTEIN. ;, mRNA sequence.
ACCESSION AI114969
NID g3515293
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 538)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP

```
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:969271
            Seq primer: custom primer used
            High quality sequence stop: 490.
FEATURES              Location/Qualifiers
     source           1. .538
                      /organism="Mus musculus"
                      /strain="C57BL"
                      /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                      Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                      with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTT];
                      double-stranded cDNA was ligated to a  DraIII adaptor
                      [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                      sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                      CACCATGTG). XhoI should be used to isolate the cDNA
                      insert.  Size selection was performed to exclude fragments
                      <1.5kb.  Library constructed by Dr. Sumio Sugano
                      (University of Tokyo Institute of Medical Science).
                      Custom primers  for sequencing: 5' end primer
                      CTTCTGCTCTAAAAGCTGCG and 3' end primer
                      CGACCTGCAGCTCGAGCACA."
                      /db_xref="taxon:10090"
                      /clone="1884947"
                      /clone_lib="Sugano mouse embryo mewa"
                      /dev_stage="embryo, 14 dpc"
                      /lab_host="DH10B"
BASE COUNT       91 a    149 c    190 g    108 t
ORIGIN


  Query Match             4.5%;  Score 110;  DB 30;  Length 538;
  Best Local Similarity   54.8%;  Pred. No. 2.4e-15;
  Matches  218;  Conservative    0;  Mismatches  180;  Indels    0;  Gaps    0;

Qy   1641 CTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAA 1700
          ||   ||  |||| |||||  ||||| | ||| |  | |||||| || | | | |  |
Db    525 CTGTTGCTGCCACTGCTGCTGCCACCACTGCTGCCCCCACTGCTACCACAGCAGCTGCTG 466

Qy   1701 CCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTG 1760
          |||| |  || |   |  |||||||  |  |  | || ||| |  | |||||||| | | |
Db    465 CCACCGCTACTGCCACCGCTACTGCCACCACTGCTGCTGCCACAGCAACTGCTGCCACCG 406

Qy   1761 AAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTG 1820
               |  ||| |  ||  | ||  |  | |||||| |  |  ||||  |||||||||
Db    405 CTGCTGCTGCCACAGCAACTGCTGCCACCGCTGCTGCCACAGCAACTGCTGCCACCGCTG 346

Qy   1821 CCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCAC 1880
          |  ||||||| ||  | ||  |||||| | ||| ||   |  | || |   |  |  ||
Db    345 CTGCCACCGCTGCTGCCACCGCTGCTGCCACAGCAACTGCTGCCACCGCTGCTGCCACAG 286

Qy   1881 CTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCA 1940
          | |||||||||| || || |  |||||  |      |    ||  | |  || |    |
Db    285 CAACTGCTGCCACCGCTGCTACAGCATCCAGAGCTGTGACGTCGGCGACGGAGAGAGAAT 226

Qy   1941 CTGCTGCCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTG 2000
          | |            |      || | ||| |||   ||| ||   | | | | |||   |
Db    225 CGGGGCTTGTGGTGGCTCAGACAGCAGCCACCACATCCAGAGCTGCAGCAGCCCCCGGAG 166

Qy   2001 CTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCA 2038
          |||   |||| ||||| |||   |||   ||    |||
Db    165 CTGGAACCACAGCAGCCCCCAGAACCCAGGCTACAGCA 128


RESULT   2
AI131654
LOCUS       AI131654      814 bp    mRNA            EST       14-SEP-1998
DEFINITION  ui41b06.x1 Sugano mouse embryo mewa Mus musculus cDNA clone 1884947
            3' similar to TR:O14633 O14633 SKIN-SPECIFIC PROTEIN. ;, mRNA
            sequence.
ACCESSION   AI131654
NID         g3601670
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 814)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:969271
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 375.
FEATURES              Location/Qualifiers
     source           1. .814
                      /organism="Mus musculus"
                      /strain="C57BL"
                      /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                      Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                      with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTT];
                      double-stranded cDNA was ligated to a  DraIII adaptor
                      [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                      sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                      CACCATGTG). XhoI should be used to isolate the cDNA
                      insert.  Size selection was performed to exclude fragments
                      <1.5kb.  Library constructed by Dr. Sumio Sugano
                      (University of Tokyo Institute of Medical Science).
                      Custom primers  for sequencing: 5' end primer
                      CTTCTGCTCTAAAAGCTGCG and 3' end primer
                      CGACCTGCAGCTCGAGCACA."
                      /db_xref="taxon:10090"
                      /clone="1884947"
                      /clone_lib="Sugano mouse embryo mewa"
                      /dev_stage="embryo, 14 dpc"
                      /lab_host="DH10B"
BASE COUNT      184 a    245 c    211 g    173 t      1 others
ORIGIN


  Query Match             4.4%;  Score 107.8;  DB 30;  Length 814;
  Best Local Similarity   54.7%;  Pred. No. 7.9e-15;
  Matches  214;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy   1641 CTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAA 1700
          ||   ||  |||| |||||  ||||| | ||| |  | |||||| || | | | |  |
Db    337 CTGTTGCTGCCACTGCTGCTGCCACCACTGCTGCCCCCACTGCTACCACAGCAGCTGCTG 396

Qy   1701 CCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTG 1760
          |||| |  || |   |  |||||||  |  |  | || ||| |  | |||||||| | | |
Db    397 CCACCGCTACTGCCACCGTTACTGCCACCACTGCTGCTGCCACAGCAACTGCTGCCACCG 456

Qy   1761 AAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTG 1820
               |  ||| |  ||  | ||  |  | |||||| |  |  ||||  |||||||||
Db    457 CTGCTGCTGCCACAGCAACTGCTGCCACCGCTGCTGCCACAGCAACTGCTGCCACCGCTG 516

Qy   1821 CCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCAC 1880
```

```
Db    517 CTGCCACCGCTGCTGCCACCGCTGCTGCCACAGCAACTGCTGCCACCGCTGCTGCCACAG 576

Qy   1881 CTACTGCTGCCCCCCACTACCGCAGCACCTGCCCCCCAACACCACAGTCACTGTACCACCCA 1940

Db    577 CAACTGCTGCCACCGCTGCTACAGCATCCAGAGCTGTGACGTCGGCGACGGAGAGAGAAT 636

Qy   1941 CTGCTGCCCCCTACTACCGCAGCACCTGCCCCCCAACACCACAGTCACTGTACCACCCACTG 2000

Db    637 CGGGGCTTGTGGTGGCTCAGACAGCAGCCACCACATCCAGAGCTGCAGCAGCCCCCGGAG 696

Qy   2001 CTGCCCCCACTGCAGCTCCCCCTACCGTCGC 2031

Db    697 CTGGAACCACAGCAGCCCCCAGAACCCAGGC 727


RESULT   3
W29419/c
LOCUS       W29419       1075 bp    mRNA            EST       11-SEP-1996
DEFINITION  mb98h11.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 337509
            5', mRNA sequence.
ACCESSION   W29419
NID         g1309566
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 1075)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:218909
            Seq primer: ETPrimer
            High quality sequence stop: 64.
FEATURES             Location/Qualifiers
     source          1. .1075
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
                     /db_xref="taxon:10090"
                     /clone="337509"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
     mRNA            <1. .>1075
BASE COUNT       37 a     82 c    789 g    159 t      8 others
ORIGIN
```

```
  Query Match             3.9%;  Score 96.8;  DB 14;  Length 1075;
  Best Local Similarity   48.8%;  Pred. No. 2.5e-12;
  Matches  315;  Conservative    0;  Mismatches  326;  Indels    5;  Gaps    2;

Qy   1492 CACATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGC 1551

Db    998 CACCCCTCCCACCGCCCCCACACCCACCCCCCCCCCGCCCCACGCCCCACCCCCGCACCAC 939

Qy   1552 TGCCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGC 1611

Db    938 ACACACCCCACCCCCACCCCCACCCCACGCCCCACACCCCACACACCCGCCCCCCGCCAC 879

Qy   1612 CCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGC 1671

Db    878 CACCCCCCACCCGACCCCACACCGCCCTCCCCCACACCCCACACCCACTCGCCCCACCCC 819

Qy   1672 TCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCC 1731

Db    818 ACCACCCACCCCCGCACCACAACCCCCGCCCACCCGC-CACACCCCCCCGCACCCCCCCC 760

Qy   1732 CACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTAC 1791

Db    759 CACATCCACCCACACCCCACCCCTCACCGCAGCCACCCTCACCCCTCGTCCACCCTCCCC 700

Qy   1792 CGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCC 1851

Db    699 CCCAACCCCCACCGCAGCTCCCCGCGCGAGCCCCCTCACACCC----CAGCCACTCCACG 644

Qy   1852 AGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGC 1911

Db    643 ACACACCCCCCACCCACACCAACCCCCCCCAGACCCCACCGCCCCCCGCGCCACACACCC 584

Qy   1912 CCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCC 1971

Db    583 CCCAACACCCCCCCCGCCTCCCCCGCCCCACGCCACCCCACCGCCCCCGCCACCCTCCCC 524

Qy   1972 CAACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGC 2031

Db    523 CCCCACCCCCCCCCCGCCCCCACCACCCCCGCCACACCCCCGCAACCCCCCCACCCCCCA 464

Qy   2032 ACATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACC 2091

Db    463 CCACGCCACCCCACCCGCACACCCACCNCNNNCGCCCGGCCCACGCCCCCCCCCACCCCGC 404

Qy   2092 TGAAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAAC 2137

Db    403 ACCGGCACCCGCCACCCCCCCCCCCCCACACCTCCACCCCACCAAC 358


RESULT   4
W34662/c
LOCUS       W34662       967 bp    mRNA            EST       11-SEP-1996
DEFINITION  mc32c11.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 350228
            5', mRNA sequence.
ACCESSION   W34662
NID         g1316708
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE   1  (bases 1 to 967)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
```

```
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:222028
            Seq primer: ETPrimer
            High quality sequence stop: 320.
FEATURES              Location/Qualifiers
     source           1. .967
                      /organism="Mus musculus"
                      /note="Vector: pT7T3D (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was size selected, ligated to Eco RI
                      adapters (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of a modified pT7T3 vector
                      (Pharmacia). Library went through one round of
                      normalization to a Cot = 5. Library constructed by Bento
                      Soares and M.Fatima Bonaldo. RNA was kindly provided by
                      Dr. Minoru Ko (Wayne State University)."
                      /db_xref="taxon:10090"
                      /clone="350228"
                      /clone_lib="Soares mouse p3NMF19.5"
                      /dev_stage="19.5 dpc total fetus"
                      /lab_host="DH10B (ampicillin resistant)"
     mRNA             <1. .>967
BASE COUNT      202 a    228 c    296 g    238 t      3 others
ORIGIN
```

```
  Query Match             3.8%;  Score 94;  DB 14;  Length 967;
  Best Local Similarity   53.1%;  Pred. No. 1e-11;
  Matches  222;  Conservative    0;  Mismatches  195;  Indels    1;  Gaps    1;

Qy   1622 CCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACT 1681
          | | |||| |   || | ||  | ||  || |  |||   |||| |||||| || | ||
Db    521 CGTGCAACTGTCGTTGCCACAACTGCTGCCCCAACTGGTACCACAGCAGCTGCTGCCACC 462

Qy   1682 GCTGCCCCTGAATCCCCAACCACTGTCAC-AGTACCACCTACTGCTGCTCCCACCGCAGC 1740
          ||| |  |     |  |  ||||   | |  |   |  |  |  ||||| ||||||  ||
Db    461 GCTACTGCCACCGCTACTGCCACCAACTCGTGCTGCCACAGCAACTGCTGCCACCGGTGC 402

Qy   1741 CCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCC 1800
            || | || ||     |||  | | |      |  | |  | ||  |  |  ||||  |
Db    401 TGCTGCCACAGCAACTGCTGCCACCGCTGCTGCCACCGCTGCTGCCACAGCAACTGCTGC 342

Qy   1801 CACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCC 1860
          |||||||||  ||||||||||  |||||||  ||  | ||  |||||| | ||| ||   |
Db    341 CACCGCTGCTGCCACCGCTGCTGCCACCGCTGCTGCCACCGCTGCTGCCACAGCAACTGC 282

Qy   1861 AACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACAC 1920
            | || |   |  |  || | |||||||||  || || |  |||||  |      |    ||
Db    281 TGCCACCGCTGCTGCCACAGCAACTGCTGCCACCGCTGCTACAGCATCCAGAGCTGTGAC 222

Qy   1921 CACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACACCAC 1980
            | |  || |    |    | |              |     || | ||| |||   |||
Db    221 GTCGGCGACGGAGAGAGAATCGGGGCTTGTGGTGGCTCAGACAGCAGCCACCACATCCAG 162

Qy   1981 AGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCA 2038
          ||   | | | |||   ||||   |||| ||||| |||   |||   ||    |||
Db    161 AGCTGCAGCAGCCCCCGGAGCTGGAACCACAGCAGCCCCCAGAACCCAGGCTACAGCA 104
```

```
RESULT   5
AU034820
LOCUS       AU034820      616 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLE543, mRNA sequence.
ACCESSION   AU034820
NID         g3800244
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 616)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES              Location/Qualifiers
     source           1. .616
                      /organism="Dictyostelium discoideum"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone_lib="SL"
                      /dev_stage="slug"
BASE COUNT      185 a    194 c     55 g    181 t      1 others
ORIGIN
```

```
  Query Match             3.7%;  Score 90.4;  DB 31;  Length 616;
  Best Local Similarity   50.3%;  Pred. No. 6.4e-11;
  Matches  220;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;

Qy   1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553
          | ||  ||||| |  |  || |  |  || | |||||     |  || |  |  | |||
Db     44 CTTCTGCTGCTTCATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTG 103

Qy   1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613
          |  |||   |  | || ||  |    |   ||    |  | ||| |  || |||||  |
Db    104 CTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCAT 163

Qy   1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673
          ||  | |  |  || |  ||  | |  |  |  |  || |  |  |  | ||| |
Db    164 CAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCAT 223

Qy   1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733
          | |  ||||| |  ||  |||  || |  ||    | | | || || |  ||||  |
Db    224 CAAGCTCTGCTCCATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCAT 283

Qy   1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793
          | |||||  ||||  ||||     |       | |   |||  ||  || | | ||
Db    284 CTGCAGCAACTACAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTA 343

Qy   1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
          |  || | ||| | ||  | ||| | ||| | ||| ||||  | || || ||||   |
Db    344 CAACCGCAACCACAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCA 403

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |    |||| || |   || |  || || |  || |  || | ||| ||||  |  ||
Db    404 TAGNTACAACCACAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCA 463

Qy   1914 CCAACACCACAGTCACT 1930
          |    || ||    |||
Db    464 CAGCTACAACCACAACT 480
```

```
RESULT   6
AA589477/c
LOCUS       AA589477      452 bp    mRNA            EST       16-SEP-1997
DEFINITION  vl47h12.s1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            975431 3', mRNA sequence.
ACCESSION   AA589477
```

```
NID         g2402857
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:556159
            Possible reversed clone: polyT not found
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:10090"
                     /clone="975431"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
     mRNA            complement(<1. .>452)
BASE COUNT       32 a     93 c    193 g    134 t
ORIGIN
```

```
  Query Match             3.7%;  Score 90.2;  DB 22;  Length 452;
  Best Local Similarity   54.3%;  Pred. No. 6.9e-11;
  Matches  182;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy   1515 CTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTCCA 1574
          ||  | |  |  | ||||   |  | ||| || | || ||| ||   ||| || || |
Db    346 CTTTTGCTGCAACAGCCCTTCCCACAGCCACAGCCGCAGCCACAGGAGCCACAGCCACAG 287

Qy   1575 CCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAA 1634
            | ||   || | ||  || || | |||        |    |  || |  | || |||
Db    286 GAGCGGCAGCAGCAGCGGCGGCAGCAGCAGACTACCACTGGCTTGCAGCAGCCACAGCAG 227

Qy   1635 TTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAAT 1694
             |  |  |  || | || ||  |  ||||| ||||  | || ||  | ||| | | |
Db    226 CCACAGCAGCCACCACAACAGCCACAGCCACCACAGCCACCACAACCACAGCCACCGCAG 167

Qy   1695 CCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTG 1754
          || || |||| |    |   | |  | || ||  |  ||||| ||||| |  | || || | |
Db    166 CCACAGCCACCGCAGCCACAGCCACCACAGCCACAGCCACCACAGCCACAGCCACCGCAG 107

Qy   1755 TCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCA 1814
           | | |  | | ||  |||   |||  |||| |  | ||||||  |  ||||  |  |||
Db    106 CCACAGCCACCACAGCCACAACCACCACAGCCACCGCAGCTGCCACAGCCGCCACAGCCA 47
```

```
Qy   1815 CCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTC 1849
          || |  || |||| ||  |  | ||   || ||||
Db     46 CCACAACCTCCACAGCCACAGCAACCCATGTTGTC 12
```

```
RESULT   7
AA798859/c
LOCUS       AA798859      514 bp    mRNA            EST       10-FEB-1998
DEFINITION  vv94a07.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            1230036 5', mRNA sequence.
ACCESSION   AA798859
NID         g2861814
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:655628
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .514
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:10090"
                     /clone="1230036"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
BASE COUNT       41 a    111 c    207 g    155 t
ORIGIN
```

```
  Query Match             3.7%;  Score 90.2;  DB 26;  Length 514;
  Best Local Similarity   54.3%;  Pred. No. 7e-11;
  Matches  182;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy   1515 CTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTCCA 1574
          ||  | |  |  | ||||   |  | ||| || | || ||| ||   ||| || || |
Db    352 CTTTTGCTGCAACAGCCCTTCCCACAGCCACAGCCGCAGCCACAGGAGCCACAGCCACAG 293

Qy   1575 CCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAA 1634
            | ||   || | ||  || || | |||        |    |  || |  | || |||
Db    292 GAGCGGCAGCAGCAGCGGCGGCAGCAGCAGACTACCACTGGCTTGCAGCAGCCACAGCAG 233

Qy   1635 TTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAAT 1694
```

```
Db    232 CCACAGCAGCCACCACAACAGCCACAGCCACCACAGCCACCACAACCACAGCCACCGCAG 173

Qy   1695 CCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTG 1754
Db    172 CCACAGCCACCGCAGCCACAGCCACCACAGCCACAGCCACCACAGCCACAGCCACCGCAG 113

Qy   1755 TCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCA 1814
Db    112 CCACAGCCACCACAGCCACAACCACCACAGCCACCGCAGCTGCCACAGCCGCCACAGCCA 53

Qy   1815 CCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTC 1849
Db     52 CCACAACCTCCACAGCCACAGCAACCCATGTTGTC 18


RESULT   8
AU034104
LOCUS       AU034104      672 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLB746, mRNA sequence.
ACCESSION   AU034104
NID         g3799528
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 672)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 672)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT - 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      214 a    207 c     60 g    190 t      1 others
ORIGIN


  Query Match             3.7%;  Score 90;  DB 31;  Length 672;
  Best Local Similarity   50.1%;  Pred. No. 8e-11;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553
Db     78 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 137

Qy   1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613
Db    138 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 197

Qy   1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673
Db    198 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 257

Qy   1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733
Db    258 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 317

Qy   1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793
```

```
Db    318 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 377

Qy   1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
Db    378 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 437

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
Db    438 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 497

Qy   1914 CCAACACCACAGTCACTGTACCA 1936
Db    498 CANCTACCTCTTAAATTTAATCA 520


RESULT   9
AA664125/c
LOCUS       AA664125      482 bp    mRNA            EST       12-NOV-1997
DEFINITION  ac06c04.s1 Stratagene lung (#937210) Homo sapiens cDNA clone 855654
            3' similar to contains element MSR1 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AA664125
NID         g2618116
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Homo sapiens"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. normal lung.  Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
                     /db_xref="taxon:9606"
                     /clone="855654"
                     /clone_lib="Stratagene lung (#937210)"
                     /sex="male"
                     /dev_stage="72 years"
                     /lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT       51 a    113 c    176 g    142 t
ORIGIN


  Query Match             3.6%;  Score 89.6;  DB 24;  Length 482;
  Best Local Similarity   50.0%;  Pred. No. 9.4e-11;
  Matches  224;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy   1497 CGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCC 1556
Db    459 CTACAGCAGCAGCAGCAGCAGCAGCAACAGCAGCAGTTCCAGGCTCAGCAGAGTGCCATG 400

Qy   1557 CAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAG 1616
```

```
          || | || |  ||   ||     |   || | ||| |  | || ||| | |||||  |||
Db    399 CAGCAGCAGTTCCAAGCAGTAGTGCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCAG 340

Qy   1617 TCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTA 1676
          || |  | |    |||| |  |  |     |  |  | |   |  | |||||  |
Db    339 CAGCAGCATCTAATTAAATTGCATCATCAAAATCAGCAACAGATACAGCAGCAGCAACAG 280

Qy   1677 CTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCG 1736
          |  |||| ||   |    |  |  |  ||   |||  | || |  | || ||  | | ||
Db    279 CAGCTGCAGCGAATAGCACAGCTGCAGCTCCAACAACAGCAACAGCAGCAGCAGCAGCAG 220

Qy   1737 CAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTG 1796
          ||||  |  |  | || | |  || |  ||||  |||     || ||||  |    || |
Db    219 CAGCAGCAGCAGCAGGCTTTGCAGGCCCAGCCACCAATTCAGCAGCCACCGATGCAG 160

Qy   1797 CCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAA 1856
          |  |||| ||  || ||  | | | | |  || ||||  || |  | | ||   || | |
Db    159 CAGCCACAGCCTCCGCCCTCCCAGGCTCTGCCCCAGCAGCTGCAGCAGATGCATCACACA 100

Qy   1857 TCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCA 1916
          || |  | | | |  |  || |  | || ||| ||| |  | ||| |||  ||  ||
Db     99 CAGCACCACCAGCCGCCACCACAGCCCCAGCAGCCTCCAGTTGCTCAGAACCAACCATCA 40

Qy   1917 ACACCACAGTCACTGTACCACCCACTGC 1944
          ||  | | ||| ||  ||  | | ||
Db     39 CAACTCCCGCCACAGTCGCAGACCCAGC 12


RESULT  10
AU033402
LOCUS       AU033402      690 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLA726, mRNA sequence.
ACCESSION   AU033402
NID         g3798826
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 690)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      212 a    213 c     67 g    198 t
ORIGIN


  Query Match             3.6%;  Score 89.4;  DB 31;  Length 690;
  Best Local Similarity   50.1%;  Pred. No. 1.1e-10;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553
          ||||  |  |  |  |  | |||||     |  || |  |  |  |||| |||   ||
Db    100 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 159

Qy   1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613
          | ||| || |  |  |  | | |     |||| || || |  || |  |||   |  |||

Db    160 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 219

Qy   1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673
          ||    || || || |  |||   || | ||  |  |  | ||  |  | | | | ||||
Db    220 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 279

Qy   1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733
          |  || |  ||||  |    ||  |  |  |      ||    || |  |||| ||  | |
Db    280 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 339

Qy   1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793
          | || ||  |||| | |||      | |       | |  ||||  ||  || |  | |||
Db    340 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 399

Qy   1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
          | ||  | ||| | ||| | ||| | ||  | ||| ||||  | || |  |||    |
Db    400 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 459

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
           |    |||||||     || |  || | |  |  | |  | || ||| ||||  |  ||
Db    460 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 519

Qy   1914 CCAACACCACAGTCACTGTACCA 1936
          |    ||| |    | |  | ||
Db    520 CAGCTACCTCTTAAATTTAATCA 542


RESULT  11
AU033450
LOCUS       AU033450      566 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLA841, mRNA sequence.
ACCESSION   AU033450
NID         g3798874
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 566)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 566)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .566
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      182 a    179 c     46 g    159 t
ORIGIN


  Query Match             3.6%;  Score 89.4;  DB 31;  Length 566;
  Best Local Similarity   50.1%;  Pred. No. 1.1e-10;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553
          ||||  |  |  |  |  | |||||     |  || |  |  |  |||| |||   ||
Db      6 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 65

Qy   1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613
          | ||| || |  |  |  | | |     |||| || || |  || |  |||   |  |||
Db     66 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 125
```

```
Qy   1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673
          ||    || || || |  |||   || | ||  |  |  | ||  |  | | | | ||||
Db    126 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 185

Qy   1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733
          |  || |  |||| |    ||  |  |  |      ||    || |  |||| ||  | |
Db    186 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 245

Qy   1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793
          | || ||  |||| | |||     | |      | |  ||||  ||  || | | |||
Db    246 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 305

Qy   1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
          | ||  | ||| | ||| | ||| | ||  | ||| |||  | || |  |||    |
Db    306 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 365

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |    |||||||     || |  || |  |  |  | || ||| ||||  |  ||
Db    366 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 425

Qy   1914 CCAACACCACAGTCACTGTACCA 1936
          |    ||| |    | |  | ||
Db    426 CAGCTACCTCTTAAATTTAATCA 448


RESULT  12
AU033594
LOCUS       AU033594      574 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLB210, mRNA sequence.
ACCESSION   AU033594
NID         g3799018
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 574)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      182 a    182 c     47 g    162 t      1 others
ORIGIN


  Query Match             3.6%;  Score 89.4;  DB 31;  Length 574;
  Best Local Similarity   50.1%;  Pred. No. 1.1e-10;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553
          ||||  |  |  |  |  | |||||     |  || |  |  |  ||||  |||   ||
Db     14 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 73

Qy   1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613
          | ||| || |  |  |  | | |     |||| || || |  || |  |||   |  |||
Db     74 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 133

Qy   1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673
          ||    || || || |  |||   || | ||  |  |  | ||  |  | | | | ||||
Db    134 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 193

Qy   1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733
          |  || |  |||| |    ||  |  |  |      ||    || |  |||| ||  | |
Db    194 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 253

Qy   1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793
          | || ||  |||| | |||     | |      | |  ||||  ||  || | | |||
Db    254 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 313

Qy   1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
          | ||  | ||| | ||| | ||| | ||  | ||| |||  | || |  |||    |
Db    314 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 373

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |    |||||||     || |  || |  |  |  | || ||| ||||  |  ||
Db    374 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 433

Qy   1914 CCAACACCACAGTCACTGTACCA 1936
          |    ||| |    | |  | ||
Db    434 CAGCTACCTCTTAAATTTAATCA 456


RESULT  13
AU033617
LOCUS       AU033617      664 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLB234, mRNA sequence.
ACCESSION   AU033617
NID         g3799041
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 664)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .664
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      193 a    213 c     66 g    192 t
ORIGIN


  Query Match             3.6%;  Score 89.4;  DB 31;  Length 664;
  Best Local Similarity   50.1%;  Pred. No. 1.1e-10;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553
          ||||  |  |  |  |  | |||||     |  || |  |  |  ||||  |||   ||
Db    103 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 162

Qy   1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613
          | ||| || |  |  |  | | |     |||| || || |  || |  |||   |  |||
Db    163 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 222

Qy   1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673
```

```
Db    223 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 282

Qy   1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733
Db    283 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 342

Qy   1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793
Db    343 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 402

Qy   1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
Db    403 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 462

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
Db    463 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 522

Qy   1914 CCAACACCACAGTCACTGTACCA 1936
Db    523 CAGCTACCTCTTAAATTTAATCA 545


RESULT  14
AU033659
LOCUS       AU033659      603 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLB278, mRNA sequence.
ACCESSION   AU033659
NID         g3799083
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 603)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      184 a    192 c     53 g    174 t
ORIGIN


  Query Match             3.6%;  Score 89.4;  DB 31;  Length 603;
  Best Local Similarity   50.1%;  Pred. No. 1.1e-10;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553
Db     43 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 102

Qy   1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613
Db    103 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 162

Qy   1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673
```

```
Db    163 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 222

Qy   1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733
Db    223 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 282

Qy   1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793
Db    283 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 342

Qy   1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
Db    343 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 402

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
Db    403 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 462

Qy   1914 CCAACACCACAGTCACTGTACCA 1936
Db    463 CAGCTACCTCTTAAATTTAATCA 485


RESULT  15
AU033708
LOCUS       AU033708      586 bp    mRNA            EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLB343, mRNA sequence.
ACCESSION   AU033708
NID         g3799132
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 586)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      182 a    186 c     50 g    168 t
ORIGIN


  Query Match             3.6%;  Score 89.4;  DB 31;  Length 586;
  Best Local Similarity   50.1%;  Pred. No. 1.1e-10;
  Matches  222;  Conservative    0;  Mismatches  221;  Indels    0;  Gaps    0;

Qy   1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553
Db     26 CATCTTCACCATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTT 85

Qy   1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613
Db     86 CTCCATCTTCATCTGCTGCAAGCTCTTCCCCATCTTCATCTGCATCATCAAGCTCTTCCC 145

Qy   1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673
Db    146 CATCTTCATCTGCATCATCAAGCTCTGCCCCATCTTCATCTGCATCATCAAGCTCTGCTC 205
```

```
Qy   1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733
          |  || |  |||| |    ||  |  | |     ||    || |  |||| ||  | |
Db    206 CATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCATCTGCAGCAACTA 265

Qy   1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793
          | || ||  |||| | |||     | |       | |  ||||  ||  || |  | |||
Db    266 CAGCTGCAACTACAATTGCAACTACAGCTGCAACTACAACTGCAACTACAACCGCAACCA 325

Qy   1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
          | ||  | ||| | ||| | ||| | ||  | ||| ||||  | || |  |||    |
Db    326 CAGCTACAACCACAGCCACAACCACAGCTACAACCACAGCTGCAACCATAGCTACAACCA 385

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
           |    |||||||     || |  || |  |  |  |  | || ||| |||| |  ||
Db    386 CAGCTGCAACTACAACTGCAACTACAACCGCAACCACAGCTACAACCACAGCTACAACCA 445

Qy   1914 CCAACACCACAGTCACTGTACCA 1936
          |    ||| |    | |  | ||
Db    446 CAGCTACCTCTTAAATTTAATCA 468
```

Search completed: July 18, 1999, 11:36:35
Job time: 1743 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 01:18:06 ; Search time 1514.49 Seconds
(without alignments)
3603.202 Million cell updates/sec

Title: US-09-294-663-1
Perfect score: 2455
Sequence: 1 GTAACGTTAAGTGAAAAGAA.....AAAAAAAAAAAAAAAAAAAA 2455

Scoring table: IDENTITY_NUC

Searched: 3232953 seqs, 1111407957 residues

Database : Pending_Patents_NA:*
1: /cgn2_6/ptodata/1/pna/PCTUS9_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US081A_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US081B_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US081C_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US081D_COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US082A_COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US082B_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US082C_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US083A_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US083B_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US084A_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US084B_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US084C_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US087A_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US087B_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US087C_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US088A_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US088B_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US088C_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US088D_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US089A_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US089B_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US089C_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US089D_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US089E_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US090A_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US090B_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US090C_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US091A_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US091B_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US092A_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US092B_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US092C_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US092D_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US6001_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US6002A_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US6002B_COMB.seq:*
44: /cgn2_6/ptodata/1/pna/US6003A_COMB.seq:*
45: /cgn2_6/ptodata/1/pna/US6003B_COMB.seq:*
46: /cgn2_6/ptodata/1/pna/US6004A_COMB.seq:*
47: /cgn2_6/ptodata/1/pna/US6004B_COMB.seq:*
48: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*
49: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*
50: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*
51: /cgn2_6/ptodata/1/pna/US6008A_COMB.seq:*
52: /cgn2_6/ptodata/1/pna/US6008B_COMB.seq:*
53: /cgn2_6/ptodata/1/pna/US6008C_COMB.seq:*
54: /cgn2_6/ptodata/1/pna/US6009A_COMB.seq:*
55: /cgn2_6/ptodata/1/pna/US6009B_COMB.seq:*
56: /cgn2_6/ptodata/1/pna/US6010A_COMB.seq:*
57: /cgn2_6/ptodata/1/pna/US6010B_COMB.seq:*
58: /cgn2_6/ptodata/1/pna/US6011_COMB.seq:*
59: /cgn2_6/ptodata/1/pna/US6012_COMB.seq:*
60: /cgn2_6/ptodata/1/pna/US6013_COMB.seq:*
61: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
62: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
63: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
64: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
65: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2455	100.0	2455	33	US-09-103-429-1	Sequence 1, Appli
2	2455	100.0	2455	33	US-09-103-429A-1	Sequence 1, Appli
3	2455	100.0	2455	38	US-09-294-663-1	Sequence 1, Appli
4	2354.4	95.9	2821	33	US-09-103-429-2	Sequence 2, Appli
5	2354.4	95.9	2821	33	US-09-103-429A-2	Sequence 2, Appli
6	2354.4	95.9	2821	38	US-09-294-663-2	Sequence 2, Appli
c 7	153.6	6.3	799	18	US-08-733-369A-105	Sequence 105, App
c 8	153.6	6.3	799	28	US-08-970-900-6	Sequence 6, Appli
c 9	153.4	6.2	803	18	US-08-733-369A-104	Sequence 104, App
c 10	153.4	6.2	803	28	US-08-970-900-7	Sequence 7, Appli
c 11	153.2	6.2	798	18	US-08-733-369A-112	Sequence 112, App
c 12	153.2	6.2	798	28	US-08-970-900-4	Sequence 4, Appli
13	151.6	6.2	8705	1	PCT-US99-06742-14	Sequence 14, Appl
c 14	151.6	6.2	10596	13	US-08-407-128-15	Sequence 15, Appl
c 15	151.6	6.2	10596	13	US-08-409-770-15	Sequence 15, Appl
c 16	151.6	6.2	10596	15	US-08-487-827-15	Sequence 15, Appl
c 17	151.6	6.2	10596	15	US-08-487-831-15	Sequence 15, Appl
c 18	151.6	6.2	9600	25	US-08-910-647-1	Sequence 1, Appli
c 19	151.6	6.2	2580	31	US-09-050-863-2	Sequence 2, Appli
20	151.6	6.2	5452	33	US-09-130-114-1	Sequence 1, Appli
c 21	151.6	6.2	10380	33	US-09-130-115-1	Sequence 1, Appli
c 22	133.2	5.4	947	59	US-60-129-391-160	Sequence 160, App
23	125.8	5.1	5163	1	PCT-US97-14104-1	Sequence 1, Appli
24	125.8	5.1	5318	1	PCT-US97-14104-2	Sequence 2, Appli
25	125.8	5.1	5163	18	US-08-700-651A-1	Sequence 1, Appli
26	125.8	5.1	5318	18	US-08-700-651A-2	Sequence 2, Appli
27	125.8	5.1	5163	18	US-08-700-651-1	Sequence 1, Appli
28	125.8	5.1	5318	18	US-08-700-651-2	Sequence 2, Appli
29	125.8	5.1	5318	26	US-08-928-361-3	Sequence 3, Appli
30	125.8	5.1	5163	26	US-08-928-361-4	Sequence 4, Appli
31	125.8	5.1	5318	63	US-08-928-361B-3	Sequence 3, Appli
32	125.8	5.1	5163	63	US-08-928-361B-4	Sequence 4, Appli
33	124.8	5.1	7334	26	US-08-928-361-1	Sequence 1, Appli
34	124.8	5.1	5511	26	US-08-928-361-2	Sequence 2, Appli
35	124.8	5.1	7334	63	US-08-928-361B-1	Sequence 1, Appli
36	124.8	5.1	5511	63	US-08-928-361B-2	Sequence 2, Appli
c 37	120.8	4.9	579	59	US-60-129-391-178	Sequence 178, App
c 38	114	4.6	644	65	US-60-141-135-1439	Sequence 1439, Ap
39	112.2	4.6	565	60	US-60-138-103-14338	Sequence 14338, A
c 40	110.6	4.5	470	59	US-60-128-437-803	Sequence 803, App
41	110.2	4.5	1010	60	US-60-138-103-13491	Sequence 13491, A
42	108.6	4.4	32207	1	PCT-US97-01442-20	Sequence 20, Appl
43	108.6	4.4	32207	1	PCT-US97-13346-20	Sequence 20, Appl
44	108.6	4.4	12754	16	US-08-592-963B-58	Sequence 58, Appl
45	108.6	4.4	12753	16	US-08-592-963B-59	Sequence 59, Appl

ALIGNMENTS

RESULT 1

```
US-09-103-429-1
; Sequence 1, Application US/09103429
;  GENERAL INFORMATION:
;   . APPLICANT:  Granados, Robert R
;     APPLICANT:  Wang, Ping
;     TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;     TITLE OF INVENTION:  cDNA and Related Products and Methods
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;       STREET:  118 North Tioga
;       CITY:  Ithaca
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  14850
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/103,429
;       FILING DATE:  24-JUN-1998
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Michaels, Christopher A
;       REGISTRATION NUMBER:  34,390
;       REFERENCE/DOCKET NUMBER:  BTI-39
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (607) 256-2000
;       TELEFAX:  (607) 256-3628
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2455 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FRAGMENT TYPE:  N-terminal
;     ORIGINAL SOURCE:
;       ORGANISM:  Trichoplusia ni
;       TISSUE TYPE:  Peritrophic Membrane
;     IMMEDIATE SOURCE:
;       CLONE:  IIM14
US-09-103-429-1
```

Query Match 100.0%; Score 2455; DB 33; Length 2455;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GTAACGTTAAGTGAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTAACGTTAAGTGAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCT 60

Qy      61 GACGGCCCTCGGGCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GACGGCCCTCGGGCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGC 120

Qy     121 TCTCCACGAGCCGCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCTCCACGAGCCGCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGA 180

Qy     181 ATACGACTGCACCAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATACGACTGCACCAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGA 240

Qy     241 CTGTGCTCCTGGTACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTGTGCTCCTGGTACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGC 300

Qy     301 CGGATGCACCCTGCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CGGATGCACCCTGCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGC 360

Qy     361 CCCAACAACCACCCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCCAACAACCACCCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGC 420

Qy     421 CCCAACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CCCAACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCC 480

Qy     481 AACTACCACTCAGGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AACTACCACTCAGGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCC 540

Qy     541 TACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAAC 600

Qy     601 TACCACTCAGGCCCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TACCACTCAGGCCCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCC 660

Qy     661 GGCCGCAACTACCCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GGCCGCAACTACCCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCC 720

Qy     721 AGGTGTTCCTGCACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AGGTGTTCCTGCACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAA 780

Qy     781 TGGTTGCCCAGCTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TGGTTGCCCAGCTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCT 840

Qy     841 CTTCTACCAGTGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTTCTACCAGTGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTT 900

Qy     901 CAACCCCTACGTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CAACCCCTACGTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAG 960

Qy     961 CCCCGCACCCCCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CCCCGCACCCCCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCT 1020

Qy    1021 CGACAATGGATGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CGACAATGGATGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTG 1080

Qy    1081 CGACAAGTATTACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCAC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CGACAAGTATTACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCAC 1140

Qy    1141 CCACTTCAGTTTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CCACTTCAGTTTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCC 1200

Qy    1201 CGGCGGCGAGAGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CGGCGGCGAGAGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCC 1260

Qy    1261 CACGGAACCCATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CACGGAACCCATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGA 1320

Qy    1321 CCACCTCCTCCCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CCACCTCCTCCCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGAC 1380

Qy    1381 TATCGCAAGACCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTC 1440
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TATCGCAAGACCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTC 1440

Qy    1441 TCCTGTGACCGCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGAC 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TCCTGTGACCGCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGAC 1500

Qy    1501 TGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAAC 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 TGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAAC 1560

Qy    1561 TGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCC 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCC 1620

Qy    1621 ACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTAC 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 ACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTAC 1680

Qy    1681 TGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGC 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 TGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGC 1740

Qy    1741 CCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCC 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 CCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCC 1800

Qy    1801 CACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCC 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 CACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCC 1860

Qy    1861 AACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACAC 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 AACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACAC 1920

Qy    1921 CACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACACCAC 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 CACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACACCAC 1980

Qy    1981 AGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACC 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 AGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACC 2040

Qy    2041 CAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGA 2100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 CAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGA 2100

Qy    2101 TGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGT 2160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 TGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGT 2160

Qy    2161 TTTGCCCTGGGCTCACGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGT 2220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 TTTGCCCTGGGCTCACGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGT 2220

Qy    2221 CCTGGTGGTTTGCTCTGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGC 2280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 CCTGGTGGTTTGCTCTGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGC 2280

Qy    2281 TTGCAACGTCGGTTGCGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCA 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 TTGCAACGTCGGTTGCGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCA 2340

Qy    2341 GGTCTTCATCCCATGGAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTT 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 GGTCTTCATCCCATGGAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTT 2400

Qy    2401 GTAAACCTACTTAAATTAATGAAGGTTTTGTTTTAAAAAAAAAAAAAAAAAAAAA 2455
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 GTAAACCTACTTAAATTAATGAAGGTTTTGTTTTAAAAAAAAAAAAAAAAAAAAA 2455

RESULT    2
```

```
US-09-103-429A-1
; Sequence 1, Application US/09103429A
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,429A
;      FILING DATE:  24-JUN-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2455 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  Peritrophic Membrane
;    IMMEDIATE SOURCE:
;      CLONE:  IIM14
US-09-103-429A-1


  Query Match             100.0%;  Score 2455;  DB 33;  Length 2455;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2455;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTAACGTTAAGTGAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTAACGTTAAGTGAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCT 60

Qy     61 GACGGCCCTCGGGCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GACGGCCCTCGGGCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGC 120

Qy    121 TCTCCACGAGCCGCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCTCCACGAGCCGCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGA 180

Qy    181 ATACGACTGCACCAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATACGACTGCACCAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGA 240

Qy    241 CTGTGCTCCTGGTACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTGTGCTCCTGGTACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGC 300
```

```
Qy    301 CGGATGCACCCTGCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGGATGCACCCTGCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGC 360

Qy    361 CCCAACAACCACCCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCCAACAACCACCCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGC 420

Qy    421 CCCAACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCCAACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCC 480

Qy    481 AACTACCACTCAGGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AACTACCACTCAGGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCC 540

Qy    541 TACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAAC 600

Qy    601 TACCACTCAGGCCCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TACCACTCAGGCCCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCC 660

Qy    661 GGCCGCAACTACCCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GGCCGCAACTACCCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCC 720

Qy    721 AGGTGTTCCTGCACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AGGTGTTCCTGCACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAA 780

Qy    781 TGGTTGCCCAGCTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGGTTGCCCAGCTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCT 840

Qy    841 CTTCTACCAGTGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTTCTACCAGTGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTT 900

Qy    901 CAACCCCTACGTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CAACCCCTACGTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAG 960

Qy    961 CCCCGCACCCCCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CCCCGCACCCCCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCT 1020

Qy   1021 CGACAATGGATGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CGACAATGGATGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTG 1080

Qy   1081 CGACAAGTATTACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCAC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CGACAAGTATTACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCAC 1140

Qy   1141 CCACTTCAGTTTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CCACTTCAGTTTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCC 1200

Qy   1201 CGGCGGCGAGAGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CGGCGGCGAGAGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCC 1260

Qy   1261 CACGGAACCCATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CACGGAACCCATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGA 1320

Qy   1321 CCACCTCCTCCCCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CCACCTCCTCCCCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGAC 1380

Qy   1381 TATCGCAAGACCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TATCGCAAGACCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTC 1440

Qy   1441 TCCTGTGACCGCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGAC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TCCTGTGACCGCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGAC 1500

Qy   1501 TGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAAC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAAC 1560

Qy   1561 TGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCC 1620

Qy   1621 ACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTAC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTAC 1680

Qy   1681 TGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGC 1740

Qy   1741 CCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCC 1800

Qy   1801 CACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCC 1860

Qy   1861 AACTACTGTCACATCACCACCTACTGCTGCCCCCCACTACCGCAGCACCTGCCCCCAACAC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACAC 1920

Qy   1921 CACAGTCACTGTACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCCCCCAACACCAC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CACAGTCACTGTACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCCCCCAACACCAC 1980

Qy   1981 AGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACC 2040

Qy   2041 CAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGA 2100

Qy   2101 TGACATCGACCCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TGACATCGACCCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGT 2160

Qy   2161 TTTGCCCTGGGCTCACGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TTTGCCCTGGGCTCACGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGT 2220

Qy   2221 CCTGGTGGTTTGCTCTGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGC 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CCTGGTGGTTTGCTCTGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGC 2280

Qy   2281 TTGCAACGTCGGTTGCGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCA 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TTGCAACGTCGGTTGCGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCA 2340

Qy   2341 GGTCTTCATCCCATGGAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GGTCTTCATCCCATGGAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTT 2400

Qy   2401 GTAAACCTACTTAAATTAATGAAGGTTTTGTTTTAAAAAAAAAAAAAAAAAAAAA 2455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GTAAACCTACTTAAATTAATGAAGGTTTTGTTTTAAAAAAAAAAAAAAAAAAAAA 2455

RESULT   3
```

```
US-09-294-663-1
; Sequence 1, Application US/09294663.
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga Street
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/294,663
;      FILING DATE:  19-APR-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/103,429
;      FILING DATE:  24-JUN-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39-CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2455 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  Peritrophic Membrane
;    IMMEDIATE SOURCE:
;      CLONE:  IIM14
US-09-294-663-1


  Query Match             100.0%;  Score 2455;  DB 38;  Length 2455;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2455;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTAACGTTAAGTGAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTAACGTTAAGTGAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCT 60

Qy      61 GACGGCCCTCGGGCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GACGGCCCTCGGGCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGC 120

Qy     121 TCTCCACGAGCCGCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCTCCACGAGCCGCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGA 180

Qy     181 ATACGACTGCACCAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATACGACTGCACCAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGA 240

Qy     241 CTGTGCTCCTGGTACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTGTGCTCCTGGTACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGC 300

Qy     301 CGGATGCACCCTGCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CGGATGCACCCTGCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGC 360

Qy     361 CCCAACAACCACCCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCCAACAACCACCCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGC 420

Qy     421 CCCAACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CCCAACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCC 480

Qy     481 AACTACCACTCAGGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AACTACCACTCAGGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCC 540

Qy     541 TACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAAC 600

Qy     601 TACCACTCAGGCCCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TACCACTCAGGCCCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCC 660

Qy     661 GGCCGCAACTACCCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GGCCGCAACTACCCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCC 720

Qy     721 AGGTGTTCCTGCACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AGGTGTTCCTGCACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAA 780

Qy     781 TGGTTGCCCAGCTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TGGTTGCCCAGCTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCT 840

Qy     841 CTTCTACCAGTGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTTCTACCAGTGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTT 900

Qy     901 CAACCCCTACGTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CAACCCCTACGTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAG 960

Qy     961 CCCCGCACCCCCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CCCCGCACCCCCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCT 1020

Qy    1021 CGACAATGGATGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CGACAATGGATGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTG 1080

Qy    1081 CGACAAGTATTACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCAC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CGACAAGTATTACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCAC 1140

Qy    1141 CCACTTCAGTTTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CCACTTCAGTTTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCC 1200

Qy    1201 CGGCGGCGAGAGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CGGCGGCGAGAGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCC 1260

Qy    1261 CACGGAACCCATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CACGGAACCCATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGA 1320

Qy    1321 CCACCTCCTCCCCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1321 CCACCTCCTCCCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGAC 1380

Qy   1381 TATCGCAAGACCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TATCGCAAGACCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTC 1440

Qy   1441 TCCTGTGACCGCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGAC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TCCTGTGACCGCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGAC 1500

Qy   1501 TGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAAC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAAC 1560

Qy   1561 TGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCC 1620

Qy   1621 ACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTAC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTAC 1680

Qy   1681 TGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGC 1740

Qy   1741 CCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCC 1800

Qy   1801 CACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCC 1860

Qy   1861 AACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACAC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACAC 1920

Qy   1921 CACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACACCAC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACACCAC 1980

Qy   1981 AGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACC 2040

Qy   2041 CAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGA 2100

Qy   2101 TGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGT 2160

Qy   2161 TTTGCCCTGGGCTCACGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TTTGCCCTGGGCTCACGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGT 2220

Qy   2221 CCTGGTGGTTTGCTCTGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGC 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CCTGGTGGTTTGCTCTGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGC 2280

Qy   2281 TTGCAACGTCGGTTGCGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCA 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TTGCAACGTCGGTTGCGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCA 2340

Qy   2341 GGTCTTCATCCCATGGAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GGTCTTCATCCCATGGAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTT 2400

Qy   2401 GTAAACCTACTTAAATTAATGAAGGTTTTGTTTTAAAAAAAAAAAAAAAAAAAAA 2455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GTAAACCTACTTAAATTAATGAAGGTTTTGTTTTAAAAAAAAAAAAAAAAAAAAA 2455
```

```
RESULT   4
US-09-103-429-2
; Sequence 2, Application US/09103429
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,429
;      FILING DATE:  24-JUN-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2821 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      DEVELOPMENTAL STAGE:  larva
;      TISSUE TYPE:  peritrophic membrane
US-09-103-429-2


  Query Match             95.9%;  Score 2354.4;  DB 33;  Length 2821;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 2422;  Conservative    0;  Mismatches    1;  Indels   57;  Gaps    1;

Qy      13 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 72
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60

Qy      73 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120

Qy     133 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180

Qy     193 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 252
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240

Qy     253 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300
Qy    313 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 372
Db    301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360
Qy    373 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 432
Db    361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420
Qy    433 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 492
Db    421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480
Qy    493 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 552
Db    481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540
Qy    553 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 612
Db    541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600
Qy    613 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 672
Db    601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660
Qy    673 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 732
Db    661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720
Qy    733 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 792
Db    721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780
Qy    793 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 852
Db    781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840
Qy    853 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 912
Db    841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900
Qy    913 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 972
Db    901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960
Qy    973 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1032
Db    961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020
Qy   1033 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1092
Db   1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080
Qy   1093 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1152
Db   1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140
Qy   1153 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1212
Db   1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200
Qy   1213 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1272
Db   1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260
Qy   1273 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1332
Db   1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320
Qy   1333 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1392
Db   1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380
```

```
Qy   1393 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1452
Db   1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440
Qy   1453 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1512
Db   1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500
Qy   1513 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1572
Db   1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560
Qy   1573 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1632
Db   1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620
Qy   1633 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1692
Db   1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680
Qy   1693 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1752
Db   1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740
Qy   1753 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1812
Db   1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800
Qy   1813 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1872
Db   1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860
Qy   1873 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1932
Db   1861 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1920
Qy   1933 ACCACCCACTGCT---------------------------------------------- 1946
Db   1921 ACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980
Qy   1946 ----------GCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 1995
Db   1981 ACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 2040
Qy   1996 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCAACACCACAGCTGC 2055
Db   2041 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100
Qy   2056 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2115
Db   2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160
Qy   2116 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2175
Db   2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220
Qy   2176 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2235
Db   2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280
Qy   2236 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2295
Db   2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340
Qy   2296 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2355
Db   2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400
Qy   2356 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2415
Db   2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460
```

```
Qy   2416 TTAATGAAGGTTTTGTTTTA 2435
          ||||||||||||||||||||
Db   2461 TTAATGAAGGTTTTGTTTTA 2480


RESULT   5
US-09-103-429A-2
; Sequence 2, Application US/09103429A
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,429A
;      FILING DATE:  24-JUN-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2821 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      DEVELOPMENTAL STAGE:  larva
;      TISSUE TYPE:  peritrophic membrane
US-09-103-429A-2


  Query Match             95.9%;  Score 2354.4;  DB 33;  Length 2821;
  Best Local Similarity   97.7%;  Pred. No. 0;
  Matches 2422;  Conservative    0;  Mismatches    1;  Indels   57;  Gaps    1;

Qy      13 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 72
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60

Qy      73 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120

Qy     133 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180

Qy     193 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 252
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240

Qy     253 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300

Qy     313 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 372
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360

Qy     373 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420

Qy     433 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480

Qy     493 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540

Qy     553 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 612
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600

Qy     613 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660

Qy     673 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 732
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720

Qy     733 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 792
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780

Qy     793 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840

Qy     853 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900

Qy     913 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 972
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960

Qy     973 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1032
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020

Qy    1033 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080

Qy    1093 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140

Qy    1153 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200

Qy    1213 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1272
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260

Qy    1273 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1332
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320
```

```
Qy   1333 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380

Qy   1393 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440

Qy   1453 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500

Qy   1513 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560

Qy   1573 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620

Qy   1633 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680

Qy   1693 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740

Qy   1753 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800

Qy   1813 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860

Qy   1873 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1920

Qy   1933 ACCACCCACTGCT---------------------------------------------- 1946
          |||||||||||||
Db   1921 ACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980

Qy   1946 ----------GCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 1995
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 ACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 2040

Qy   1996 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCAACACCACAGCTGC 2055
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   2041 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100

Qy   2056 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2115
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160

Qy   2116 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220

Qy   2176 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280

Qy   2236 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340

Qy   2296 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400

Qy   2356 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460

Qy   2416 TTAATGAAGGTTTTGTTTTA 2435
          ||||||||||||||||||||
Db   2461 TTAATGAAGGTTTTGTTTTA 2480
```

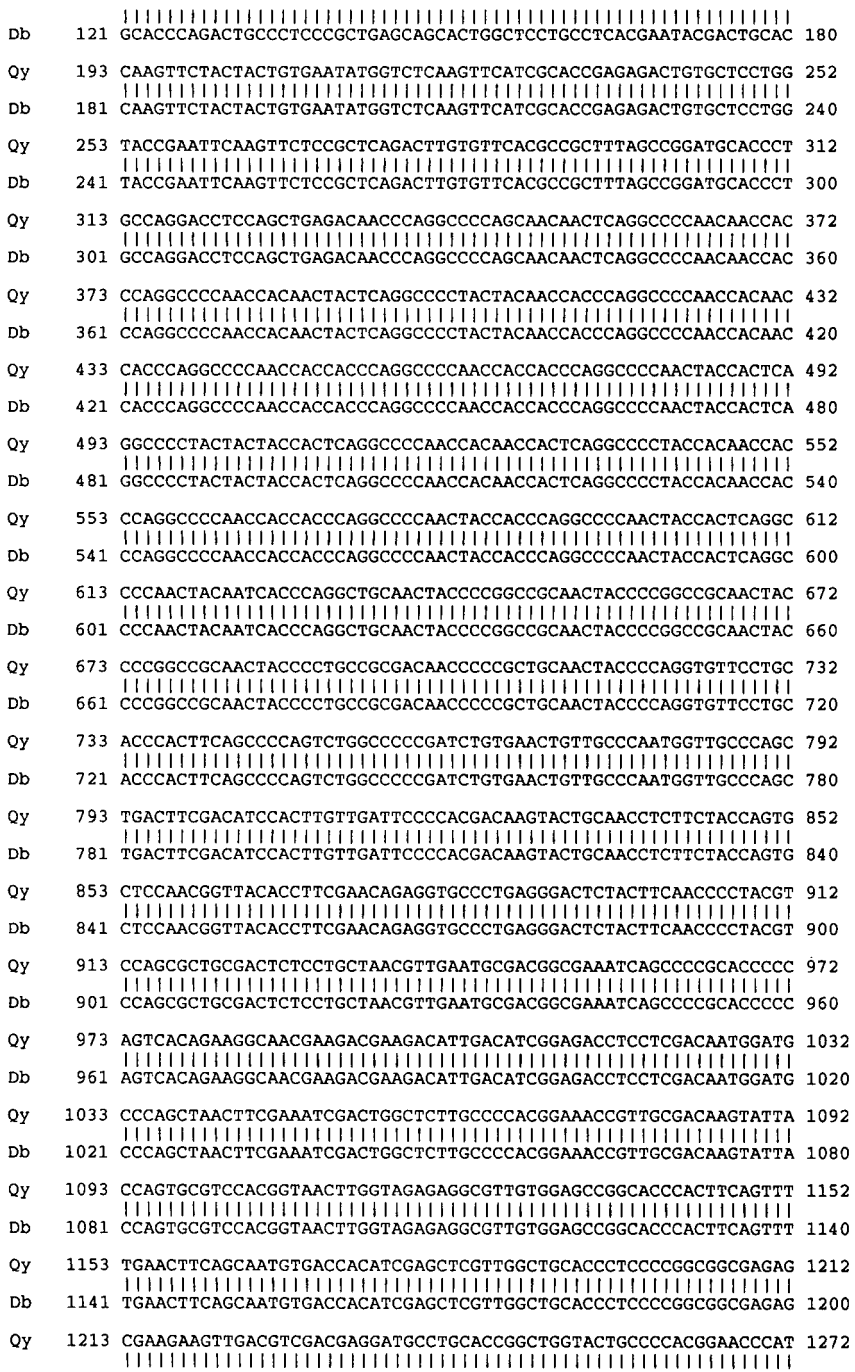
RESULT 6
US-09-294-663-2
; Sequence 2, Application US/09294663
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin
; TITLE OF INVENTION: cDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 North Tioga Street
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/294,663
; FILING DATE: 19-APR-1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/103,429
; FILING DATE: 24-JUN-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2821 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; DEVELOPMENTAL STAGE: larva
; TISSUE TYPE: peritrophic membrane
US-09-294-663-2

Query Match 95.9%; Score 2354.4; DB 38; Length 2821;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 2422; Conservative 0; Mismatches 1; Indels 57; Gaps 1;

```
Qy     13 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 72
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60

Qy     73 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120

Qy    133 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 192
```


Db 121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180
Qy 193 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 252
Db 181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240
Qy 253 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 312
Db 241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300
Qy 313 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 372
Db 301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360
Qy 373 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 432
Db 361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420
Qy 433 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 492
Db 421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480
Qy 493 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 552
Db 481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540
Qy 553 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 612
Db 541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600
Qy 613 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 672
Db 601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660
Qy 673 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 732
Db 661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720
Qy 733 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 792
Db 721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780
Qy 793 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 852
Db 781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840
Qy 853 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 912
Db 841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900
Qy 913 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 972
Db 901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960
Qy 973 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1032
Db 961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020
Qy 1033 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1092
Db 1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080
Qy 1093 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1152
Db 1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140
Qy 1153 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1212
Db 1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200
Qy 1213 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1272
Db 1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260
Qy 1273 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1332
Db 1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320
Qy 1333 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1392
Db 1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380
Qy 1393 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1452
Db 1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440
Qy 1453 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1512
Db 1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500
Qy 1513 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1572
Db 1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560
Qy 1573 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1632
Db 1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620
Qy 1633 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1692
Db 1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680
Qy 1693 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1752
Db 1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740
Qy 1753 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1812
Db 1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800
Qy 1813 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1872
Db 1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860
Qy 1873 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1932
Db 1861 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1920
Qy 1933 ACCACCCACTGCT---------- 1946
Db 1921 ACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980
Qy 1946 ----------GCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 1995
Db 1981 ACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 2040
Qy 1996 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCAACACCACAGCTGC 2055
Db 2041 CACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100
Qy 2056 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2115
Db 2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160
Qy 2116 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2175
Db 2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220
Qy 2176 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2235
Db 2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280
Qy 2236 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2295
Db 2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340

```
Qy   2296 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400

Qy   2356 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460

Qy   2416 TTAATGAAGGTTTTGTTTTA 2435
          ||||||||||||||||||||
Db   2461 TTAATGAAGGTTTTGTTTTA 2480


RESULT   7
US-08-733-369A-105/c
; Sequence 105, Application US/08733369A
;  GENERAL INFORMATION:
;    APPLICANT:  Masucci, Maria G.
;    TITLE OF INVENTION:  GLYCINE-CONTAINING SEQUENCES CONFERRING
;    TITLE OF INVENTION:  INVISIBILITY TO THE IMMUNE SYSTEM.
;    NUMBER OF SEQUENCES:  123
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kathleen M. Williams, Banner & Witcoff, Ltd.
;      STREET:  One Financial Center
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Wordperfect 6.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,369A
;      FILING DATE:  17 October, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/522,995
;      FILING DATE:  01-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/529,190
;      FILING DATE:  15-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 95013249
;      FILING DATE:  10-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB96/00876
;      FILING DATE:  10-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Kathleen M.
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  95-1391-D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-345-9100
;      TELEFAX:  617-345-9111
;  INFORMATION FOR SEQ ID NO:  105:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  799 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-733-369A-105


  Query Match             6.3%;  Score 153.6;  DB 18;  Length 799;
  Best Local Similarity   53.2%;  Pred. No. 1.2e-25;
  Matches  352;  Conservative    0;  Mismatches  304;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || | |  | |||||||  | ||| |  |  |  |||||||
Db    662 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 603

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db    602 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 543

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db    542 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 483

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||    |     | ||| ||||  ||||  || |
Db    482 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 423

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||       |     ||| | |  | || ||| | |||
Db    422 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 363

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| ||| |
Db    362 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 303

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |||  || |||   |  | || ||| ||||| |  || || | ||  | |||||||
Db    302 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 243

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |     |  |     |    ||  || || || |  || || || |  | ||||| ||
Db    242 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 183

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
          | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db    182 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 123

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
          |||  ||     || |  |||||||     || |  |   | | ||  || |  |  |
Db    122 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 63

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAAT 2153
          || |     |  ||| |||   |    |  ||  |    || || || |  | ||
Db     62 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGTCGGGGGAT 3

Qy   2154 GC 2155
          ||
Db      2 CC 1


RESULT   8
US-08-970-900-6/c
; Sequence 6, Application US/08970900
;  GENERAL INFORMATION:
;    APPLICANT:  Masucci, Maria G.
;    TITLE OF INVENTION:  FUSION PROTEINS HAVING INCREASED HALF-LIVES.
;    NUMBER OF SEQUENCES:  91
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kathleen M. Williams, Banner & Witcoff, Ltd.
;      STREET:  One Financial Center
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Wordperfect 6.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/970,900
;      FILING DATE:  14-NOV-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/030,986
;      FILING DATE:  15-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/048,945
```

```
;      FILING DATE:  25-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Kathleen M.
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  3255/59831
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-345-9100
;      TELEFAX:  617-345-9111
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  799 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-970-900-6


  Query Match             6.3%;  Score 153.6;  DB 28;  Length 799;
  Best Local Similarity   53.2%;  Pred. No. 1.2e-25;
  Matches  352;  Conservative    0;  Mismatches  304;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || |  |  |||||||  | ||| |  |  |  |||||||
Db    662 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 603

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db    602 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 543

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| ||||  | ||  | ||  |  | || ||| ||
Db    542 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 483

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||     |     | ||| ||||  ||||  || |
Db    482 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 423

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||       |     ||| | |  | || ||| | |||
Db    422 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 363

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db    362 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 303

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
              |||  || |||   |  | || ||| |||||  |  || || |  ||  | |||||||
Db    302 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 243

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |    |  |    |   ||  || || || |  || || || |  | ||||| ||
Db    242 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 183

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
          | |  | | | ||    | ||  || |||| ||  |  |  |||  || | | | | | |
Db    182 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 123

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
            ||| ||     || |  |||||||     || |  |   | | ||  || |  |  |
Db    122 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 63

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAAT 2153
            || |     |   ||| |||    |  |  ||  |   || || || |  | ||
Db     62 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGTCGGGGGAT 3

Qy   2154 GC 2155
          |
Db       2 CC 1


RESULT   9
US-08-733-369A-104/c
; Sequence 104, Application US/08733369A
;  GENERAL INFORMATION:
;    APPLICANT:  Masucci, Maria G.
;    TITLE OF INVENTION:  GLYCINE-CONTAINING SEQUENCES CONFERRING
;    TITLE OF INVENTION:  INVISIBILITY TO THE IMMUNE SYSTEM.
;    NUMBER OF SEQUENCES:  123
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kathleen M. Williams, Banner & Witcoff, Ltd.
;      STREET:  One Financial Center
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Wordperfect 6.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/733,369A
;      FILING DATE:  17 October, 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/522,995
;      FILING DATE:  01-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/529,190
;      FILING DATE:  15-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  SE 95013249
;      FILING DATE:  10-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB96/00876
;      FILING DATE:  10-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Kathleen M.
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  95-1391-D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-345-9100
;      TELEFAX:  617-345-9111
;  INFORMATION FOR SEQ ID NO:  104:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  803 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-733-369A-104


  Query Match             6.2%;  Score 153.4;  DB 18;  Length 803;
  Best Local Similarity   53.1%;  Pred. No. 1.3e-25;
  Matches  353;  Conservative    0;  Mismatches  306;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || |  |  |||||||  | ||| |  |  |  |||||||
Db    666 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 607

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db    606 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 547

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db    546 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 487

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||     |     | ||| ||||  ||||  || |
Db    486 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 427

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||       |     ||| | |  | || ||| | |||
Db    426 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 367
```

```
Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
             | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db    366 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 307

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |||  || |||    |  | || ||| |||||  |  || || | ||  | |||||||
Db    306 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 247

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 1973
           |     |  |     |    ||  || || || |  || || || |  | ||||| ||
Db    246 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 187

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
           | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db    186 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 127

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
           ||| ||     || |  |||||||     || |  |  | | ||  || | |  | |
Db    126 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 67

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAAT 2153
             || |      |   ||| |||    |    |  ||  |    || || || |  | ||
Db     66 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGTCGGGGGAT 7

Qy   2154 GCAAC 2158
             | |
Db      6 CGATC 2


RESULT  10
US-08-970-900-7/c
; Sequence 7, Application US/08970900
;  GENERAL INFORMATION:
;    APPLICANT:  Masucci, Maria G.
;    TITLE OF INVENTION:  FUSION PROTEINS HAVING INCREASED HALF-LIVES.
;    NUMBER OF SEQUENCES:  91
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kathleen M. Williams, Banner & Witcoff, Ltd.
;      STREET:  One Financial Center
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Wordperfect 6.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/970,900
;      FILING DATE:  14-NOV-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/030,986
;      FILING DATE:  15-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/048,945
;      FILING DATE:  25-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Kathleen M.
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  3255/59831
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-345-9100
;      TELEFAX:  617-345-9111
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  803 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-970-900-7

  Query Match             6.2%;  Score 153.4;  DB 28;  Length 803;
  Best Local Similarity   53.1%;  Pred. No. 1.3e-25;
  Matches  353;  Conservative    0;  Mismatches  306;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
           || |||||||  | ||| || |  |  |||||||  | |||  |  |  |  |||||||
Db    666 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 607

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
           || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db    606 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 547

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
           |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db    546 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 487

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
           || ||||||||    | ||  || ||    |    | ||| ||||  ||||  || |
Db    486 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 427

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
           | ||| || ||||    ||||       |     ||| | |  | || ||| | |||
Db    426 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 367

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
             | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db    366 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 307

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |||  || |||    |  | || ||| |||||  |  || || | ||  | |||||||
Db    306 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 247

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTACCGCAGCACCTGCCCCCA 1973
           |     |  |     |    ||  || || || |  || || || |  | ||||| ||
Db    246 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 187

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
           | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db    186 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 127

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
           ||| ||     || |  |||||||     || |  |  | | ||  || | |  | |
Db    126 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 67

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAAT 2153
             || |      |   ||| |||    |    |  ||  |    || || || |  | ||
Db     66 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGTCGGGGGAT 7

Qy   2154 GCAAC 2158
             | |
Db      6 CGATC 2


RESULT  11
US-08-733-369A-112/c
; Sequence 112, Application US/08733369A
;  GENERAL INFORMATION:
;    APPLICANT:  Masucci, Maria G.
;    TITLE OF INVENTION:  GLYCINE-CONTAINING SEQUENCES CONFERRING
;    TITLE OF INVENTION:  INVISIBILITY TO THE IMMUNE SYSTEM.
;    NUMBER OF SEQUENCES:  123
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kathleen M. Williams, Banner & Witcoff, Ltd.
;      STREET:  One Financial Center
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  Wordperfect 6.1a
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/733,369A
;       FILING DATE:  17 October, 1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/522,995
;       FILING DATE:  01-SEP-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/529,190
;       FILING DATE:  15-SEP-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  SE 95013249
;       FILING DATE:  10-APR-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/GB96/00876
;       FILING DATE:  10-APR-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Williams, Kathleen M.
;       REGISTRATION NUMBER:  34,380
;       REFERENCE/DOCKET NUMBER:  95-1391-D
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-345-9100
;       TELEFAX:  617-345-9111
;   INFORMATION FOR SEQ ID NO:  112:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  798 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
US-08-733-369A-112


  Query Match             6.2%;  Score 153.2;  DB 18;  Length 798;
  Best Local Similarity   53.2%;  Pred. No. 1.4e-25;
  Matches  351;  Conservative    0;  Mismatches  303;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || |  |  |||||||  | ||| |  |  |  |||||||
Db    661 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 602

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db    601 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 542

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db    541 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 482

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||    |     | ||| ||||  ||||  || |
Db    481 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 422

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||      |     ||| | |  | || ||| | |||
Db    421 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 362

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
             | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db    361 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 302

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
             |||  || |||    |  | || ||| |||||  |  || || | ||  | |||||||
Db    301 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 242

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |     |  |      |    ||  || || || |  || || || |  | ||||| ||
Db    241 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 182

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
           | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db    181 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 122

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
          |||  ||     || |  |||||||     || |  |   | | ||  || |  |  |
Db    121 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 62

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAAT 2153
             || |     |   ||| |||   |   |  ||  |    || || || |  | ||
Db     61 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGTCGGGGGAT 2


RESULT  12
US-08-970-900-4/c
; Sequence 4, Application US/08970900
;  GENERAL INFORMATION:
;    APPLICANT:  Masucci, Maria G.
;    TITLE OF INVENTION:  FUSION PROTEINS HAVING INCREASED HALF-LIVES.
;    NUMBER OF SEQUENCES:  91
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kathleen M. Williams, Banner & Witcoff, Ltd.
;      STREET:  One Financial Center
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Wordperfect 6.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/970,900
;      FILING DATE:  14-NOV-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/030,986
;      FILING DATE:  15-NOV-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/048,945
;      FILING DATE:  25-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Kathleen M.
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  3255/59831
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-345-9100
;      TELEFAX:  617-345-9111
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  798 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-970-900-4


  Query Match             6.2%;  Score 153.2;  DB 28;  Length 798;
  Best Local Similarity   53.2%;  Pred. No. 1.4e-25;
  Matches  351;  Conservative    0;  Mismatches  303;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || |  |  |||||||  | ||| |  |  |  |||||||
Db    661 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 602

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db    601 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 542

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db    541 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 482

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||    |     | ||| ||||  ||||  || |
```

```
Db    481 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 422

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
Db    421 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 362

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
Db    361 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 302

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
Db    301 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 242

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
Db    241 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 182

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
Db    181 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 122

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
Db    121 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 62

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAAT 2153
Db     61 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGTCGGGGGAT 2


RESULT  13
PCT-US99-06742-14
; Sequence 14, Application PC/TUS9906742
; GENERAL INFORMATION:
;  APPLICANT: Ruffner, Duane E.
;  APPLICANT: Pierce, Michael L.
;  APPLICANT: Chen, Zhidong
;  TITLE OF INVENTION: Directed Antisense Libraries
;  FILE REFERENCE: T6678.PCT
;  CURRENT APPLICATION NUMBER: PCT/US99/06742
;  CURRENT FILING DATE: 1999-03-28
;  EARLIER APPLICATION NUMBER: 60/079,792
;  EARLIER FILING DATE: 1998-03-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: Word97
; SEQ ID NO 14
;   LENGTH: 8705
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: pShuttle
PCT-US99-06742-14


  Query Match             6.2%;  Score 151.6;  DB 1;  Length 8705;
  Best Local Similarity   53.2%;  Pred. No. 8.6e-25;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
Db   7383 ctcctgcccctcctgctcctgcccctcctgcccctcctgctcctgcccctcctgccctc 7442

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
Db   7443 ctcctgctcctgcccctcctgcccctcctcctgctcctgcccctcctgcccctcctgctc 7502

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
Db   7503 ctgcccctcctgcccctcctgctcctgcccctcctgcccctcctgctcctgcccctcctg 7562

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739

Db   7563 ctcctgcccctcctgctcctgcccctcctgctcctgcccctcctgcccctcctgcccctc 7622

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
Db   7623 ctcctgctcctgcccctcctgctcctgcccctcctgcccctcctgcccctcctgctcctg 7682

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
Db   7683 ccctcctcctgctcctgccctcctgccctcctgccctcctcctgctcctgccctc 7742

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
Db   7743 ctgccctcctcctgctcctgccctcctcctgctcctgccctcctgccctcctgccc 7802

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
Db   7803 ctcctcctgctcctgccctcctgccctcctcctgctcctgccctcctcctgctcctg 7862

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
Db   7863 ccctcctgccctcctgccctcctcctgctcctgccctcctcctgctcctgccctc 7922

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
Db   7923 ctgccctcctgccctcctgccctcctcctgctcctgccctcctcctgctcctgccc 7982

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
Db   7983 ctcctgctcctgccctcccgctcctgctcctgctcctgttccaccgtgggt 8034


RESULT  14
US-08-407-128-15/c
; Sequence 15, Application US/08407128
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/407,128
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  779
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-407-128-15
```

```
  Query Match             6.2%;  Score 151.6;  DB 13;  Length 10596;
  Best Local Similarity   53.2%;  Pred. No. 9.3e-25;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || |  | |||||||  | ||| |  |  |  |||||||
Db   2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db   2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| ||||  | ||  | ||  |  | || ||| ||
Db   2747 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 2688

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||      |      | ||| |||| |||| || |
Db   2687 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||      |     ||| | |  | || ||| | |||
Db   2627 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db   2567 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |||  || |||   |  | || ||| |||||  |  || || | ||  | |||||||
Db   2507 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |     |  |     |    ||  || || || |  || || || |  | |||||  ||
Db   2447 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 2388

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
          | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db   2387 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
          |||.||    || |  |||||||    || |  |  | | ||  || |  |  |  |
Db   2327 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
            || |     |   ||| |||   |    |  ||  |    || || ||
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216


RESULT  15
US-08-409-770-15/c
; Sequence 15, Application US/08409770
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/409,770
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  755.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-409-770-15

  Query Match             6.2%;  Score 151.6;  DB 13;  Length 10596;
  Best Local Similarity   53.2%;  Pred. No. 9.3e-25;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || |  | |||||||  | ||| |  |  |  |||||||
Db   2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db   2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| ||||  | ||  | ||  |  | || ||| ||
Db   2747 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 2688

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||      |      | ||| |||| |||| || |
Db   2687 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||      |     ||| | |  | || ||| | |||
Db   2627 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db   2567 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |||  || |||   |  | || ||| |||||  |  || || | ||  | |||||||
Db   2507 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |     |  |     |    ||  || || || |  || || || |  | |||||  ||
Db   2447 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 2388

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
          | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db   2387 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
          |||.||    || |  |||||||    || |  |  | | ||  || |  |  |  |
Db   2327 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
            || |     |   ||| |||   |    |  ||  |    || || ||
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216


Search completed: July 18, 1999, 02:08:42
Job time: 3036 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 07:48:19 ; Search time 72.5 Seconds
(without alignments)
3104.358 Million cell updates/sec

Title: US-09-294-663-1
Perfect score: 2455
Sequence: 1 GTAACGTTAAGTGAAAAGAA.....AAAAAAAAAAAAAAAAAAAA 2455

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	151.6	6.2	10596	1	US-07-884-811-15	Sequence 15, Appl
c 2	151.6	6.2	10596	1	US-07-885-971-15	Sequence 15, Appl
c 3	151.6	6.2	10596	1	US-08-087-783A-15	Sequence 15, Appl
c 4	151.6	6.2	10596	1	US-08-194-088B-15	Sequence 15, Appl
c 5	151.6	6.2	10596	4	US-08-194-087-15	Sequence 15, Appl
c 6	151.6	6.2	10596	5	PCT-US93-04648-15	Sequence 15, Appl
c 7	113	4.6	2793	3	US-08-209-747-1	Sequence 1, Appli
c 8	113	4.6	2793	3	US-08-458-298-1	Sequence 1, Appli
9	108.6	4.4	32207	4	US-08-770-379-20	Sequence 20, Appl
c 10	100.6	4.1	9551	1	US-08-056-200-93	Sequence 93, Appl
c 11	99.4	4.0	6530	5	PCT-US93-03993-1	Sequence 1, Appli
12	93	3.8	1984	1	US-07-885-970A-25	Sequence 25, Appl
13	93	3.8	1985	1	US-08-298-687A-25	Sequence 25, Appl
14	93	3.8	1985	1	US-08-298-829-25	Sequence 25, Appl
15	91.4	3.7	913	1	US-08-217-327-3	Sequence 3, Appli
16	91.4	3.7	913	1	US-07-885-970A-3	Sequence 3, Appli
17	91.4	3.7	913	1	US-08-298-687A-3	Sequence 3, Appli
18	91.4	3.7	913	1	US-08-530-797-2	Sequence 2, Appli
19	91.4	3.7	913	1	US-08-298-829-3	Sequence 3, Appli
c 20	85.6	3.5	1505	1	US-07-915-246-1	Sequence 1, Appli
c 21	83.4	3.4	44377	4	US-08-804-227C-7	Sequence 7, Appli
c 22	83.4	3.4	13987	4	US-08-804-227C-13	Sequence 13, Appl
23	82.2	3.3	12001	3	US-08-458-568A-11	Sequence 11, Appl
24	81	3.3	7218	2	US-08-232-463-14	Sequence 14, Appl
c 25	77.6	3.2	2338	2	US-08-425-069-1	Sequence 1, Appli
c 26	77.2	3.1	10747	5	PCT-US93-03985-1	Sequence 1, Appli
c 27	77.2	3.1	24979	5	PCT-US93-03985-3	Sequence 3, Appli
c 28	76.4	3.1	2427	4	US-08-678-039A-39	Sequence 39, Appl
c 29	74.8	3.0	2242	2	US-08-641-627A-37	Sequence 37, Appl
c 30	71.8	2.9	43280	4	US-08-804-227C-1	Sequence 1, Appli
c 31	69.8	2.8	32207	4	US-08-770-379-20	Sequence 20, Appl
32	65.8	2.7	31571	2	US-08-323-443B-1	Sequence 1, Appli
33	65.4	2.7	3833	3	US-08-917-320-18	Sequence 18, Appl
34	65.4	2.7	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
35	64.8	2.6	786	3	US-08-403-379A-2	Sequence 2, Appli
36	64.6	2.6	2093	1	US-08-287-001A-1	Sequence 1, Appli
37	64.6	2.6	2093	5	PCT-US95-09941-1	Sequence 1, Appli
c 38	63.4	2.6	234	3	US-08-469-802B-3	Sequence 3, Appli
c 39	63.4	2.6	234	4	US-08-267-803B-3	Sequence 3, Appli
40	63.2	2.6	2032	5	PCT-US95-07721-5	Sequence 5, Appli
41	62.6	2.5	2753	4	US-07-952-853-5	Sequence 5, Appli
42	61.8	2.5	10366	2	US-08-246-982A-5	Sequence 5, Appli
43	61.8	2.5	10366	2	US-08-453-265-5	Sequence 5, Appli
44	61.8	2.5	10348	4	US-08-457-273B-41	Sequence 41, Appl
45	60.6	2.5	3624	1	US-07-951-715A-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT 1
US-07-884-811-15/c
; Sequence 15, Application US/07884811
; Patent No. 5316921
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/884,811
;      FILING DATE:  19920518
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  755.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-884-811-15


  Query Match           6.2%;  Score 151.6;  DB 1;  Length 10596;
  Best Local Similarity  53.2%;  Pred. No. 3e-23;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || |  |  |||||||  | ||| |  |  |  |||||||
Db   2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db   2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db   2747 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 2688
```

```
Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||     |     | ||| ||||  ||||  || |
Db   2687 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||      |     ||| | |  | || ||| | |||
Db   2627 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db   2567 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |||  || |||   |  | || ||| ||||| |  || || | ||  | |||||||
Db   2507 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |    | |     |   || || || || | || || || | |  | |||||
Db   2447 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTG 2388

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
            | |  | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db   2387 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
          |||  ||     || |  |||||||     || |  |  | | ||  || |  |  |
Db   2327 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
            || |     |  ||| |||  |    | ||  |    || || ||
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216
```

```
RESULT   2
US-07-885-971-15/c
; Sequence 15, Application US/07885971
; Patent No. 5328837
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,971
;      FILING DATE:  19920518
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  779
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-885-971-15

  Query Match              6.2%;  Score 151.6;  DB 1;  Length 10596;
  Best Local Similarity   53.2%;  Pred. No. 3e-23;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || | |  |||||||  | ||| |  |  |  |||||||
Db   2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db   2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db   2747 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 2688

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||     |     | ||| ||||  ||||  || |
Db   2687 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||      |     ||| | |  | || ||| | |||
Db   2627 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db   2567 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |||  || |||   |  | || ||| ||||| |  || || | ||  | |||||||
Db   2507 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |    | |     |   || || || || | || || || | |  | |||||
Db   2447 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 2388

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
            | |  | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db   2387 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
          |||  ||     || |  |||||||     || |  |  | | ||  || |  |  |
Db   2327 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
            || |     |  ||| |||  |    | ||  |    || || ||
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216
```

```
RESULT   3
US-08-087-783A-15/c
; Sequence 15, Application US/08087783A
; Patent No. 5547856
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/087,783A
;      FILING DATE:  13-Jul-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/884811
;      FILING DATE:  18-MAY-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/885971
;      FILING DATE:  18-MAY-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Marschang, Diane L.
;      REGISTRATION NUMBER:  35,600
;      REFERENCE/DOCKET NUMBER:  P0755779P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-5416
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-087-783A-15


  Query Match             6.2%;  Score 151.6;  DB 1;  Length 10596;
  Best Local Similarity   53.2%;  Pred. No. 3e-23;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || | |  | |||||||  | ||| |  |  |  |||||||
Db   2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db   2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db   2747 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 2688

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||    |     | ||| ||||  ||||  || |
Db   2687 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||      |     ||| | |  | || ||| | |||
Db   2627 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db   2567 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
             |||  || |||   |  | || ||| |||||  |  || || | ||  | |||||||
Db   2507 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |    |  |    |   ||  || || || |  || || || |  | |||||  ||
Db   2447 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 2388

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
             | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | |
Db   2387 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
           ||| ||     || |  |||||||     || |  |   | | ||  || | |  |
Db   2327 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
               || |     |   ||| |||   |    |  ||  |    || || ||
```

```
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216


RESULT   4
US-08-194-088B-15/c
; Sequence 15, Application US/08194088B
; Patent No. 5580963
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,088B
;      FILING DATE:  09-FEB-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/884811
;      FILING DATE:  18-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gallegos, R. Thomas
;      REGISTRATION NUMBER:  32,692
;      REFERENCE/DOCKET NUMBER:  755D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-2614
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-194-088B-15


  Query Match             6.2%;  Score 151.6;  DB 1;  Length 10596;
  Best Local Similarity   53.2%;  Pred. No. 3e-23;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || | |  | |||||||  | ||| |  |  |  |||||||
Db   2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db   2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db   2747 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 2688

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||    |     | ||| ||||  ||||  || |
Db   2687 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||      |     ||| | |  | || ||| | |||
Db   2627 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
```

```
Db   2567 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
Db   2507 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
Db   2447 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 2388

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
Db   2387 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
Db   2327 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216


RESULT   5
US-08-194-087-15/c
; Sequence 15, Application US/08194087
; Patent No. 5879910
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,087
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  779
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-194-087-15


  Query Match             6.2%;  Score 151.6;  DB 4;  Length 10596;
  Best Local Similarity   53.2%;  Pred. No. 3e-23;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
```

```
Db   2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
Db   2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
Db   2747 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 2688

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
Db   2687 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
Db   2627 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
Db   2567 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
Db   2507 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
Db   2447 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 2388

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
Db   2387 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
Db   2327 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216


RESULT   6
PCT-US93-04648-15/c
; Sequence 15, Application PC/TUS9304648
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc., Godowski, Paul J., Lokker, Natalie A., Mark, Melani
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/04648
;      FILING DATE:  19930517
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/884811
;      FILING DATE:  18-MAY-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/885971
;      FILING DATE:  18-MAY-92
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  755,779P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US93-04648-15


  Query Match             6.2%;  Score 151.6;  DB 5;  Length 10596;
  Best Local Similarity   53.2%;  Pred. No. 3e-23;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || ||||||| | ||| || | | | ||||||| | ||| | | | |||||||
Db   2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| | | ||  || |  |||| | || |  |
Db   2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  | | ||   |||| |||| ||||| | ||  | ||  |  |'|| ||| ||
Db   2747 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 2688

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || |||||||    | ||  || ||    |     | ||| |||| |||| || |
Db   2687 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||      |     ||| | |  | || ||| | |||
Db   2627 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db   2567 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |||  || |||   |  | || ||| ||||| |  || || | ||  | |||||||
Db   2507 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |     |  |     |    ||  || || || |  || || || |  | ||||| ||
Db   2447 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 2388

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
          | |  | | | ||    | ||  || |||| ||  |  |  |||| || | | | | |
Db   2387 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
            ||| ||     || |  |||||||     || |  |   | | ||  || | |  |
Db   2327 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
            || |     |   ||| |||   |   | ||  |    || || ||
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216


RESULT   7
US-08-209-747-1/c
; Sequence 1, Application US/08209747
; Patent No. 5733771
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/209,747
;      FILING DATE:  14-MAR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-209-747-1


  Query Match             4.6%;  Score 113;  DB 3;  Length 2793;
  Best Local Similarity   50.9%;  Pred. No. 2.3e-15;
  Matches  269;  Conservative    0;  Mismatches  260;  Indels    0;  Gaps    0;

Qy   1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553
          || || || | || ||  |||| |||||| |  | || ||      |    ||| | ||
Db   2120 CAGCGGCTCCAGCTCCTGCAGCGGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGC 2061

Qy   1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613
          | ||   | | |||||  |  | |  |   |||| ||  |  | || ||| |    | |
Db   2060 CTCCGGATCCTGCACCAGCAGCTGCAGCAGCACCAGCTCCTGCTCCGGCACCGTATCCAC 2001

Qy   1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673
          |    |||||     |  ||  |||| |||| || ||  | || || ||  |  |||| |
Db   2000 CTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCAC 1941

Qy   1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733
          |  |  || |||| |    | | | |         | |||  || ||  || ||||||| |
Db   1940 CAGCGGCTCCTGCACGACCACCTCTACTGTAACCCGCAGCGCCTCCAGCTCCTGCTCCAG 1881

Qy   1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793
          | || ||| |  |  | ||| |  ||      ||      |   |  | || ||    |
Db   1880 CAGCTGCCGCAGCACCAGCTCTTGCTCCGGCACCGTATCCACCCTGACCACCGTAGCCTC 1821

Qy   1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
          | || ||  |  |||| ||  | || || ||  |  |||| ||| |  | |||    |
Db   1820 CAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCACCTGCAGCAGCTCCAGCTC 1761

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
```

```
Db   1760 CAGCACCAGCACCTCTACCGTAACCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCG 1701

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973

Db   1700 CACCAGCTCCTGCTCCGGCACCATAACCGCCTTGTCCACCATATCCAGCACCACCGGTGC 1641

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCC 2022

Db   1640 CTGCACCTGCCCCAGCAGCGGCACCAGCTCCAGCACCAGCACCTCTACC 1592


RESULT   8
US-08-458-298-1/c
; Sequence 1, Application US/08458298
; Patent No. 5756677
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,298
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,747
;      FILING DATE:  14-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-458-298-1


  Query Match             4.6%;  Score 113;  DB 3;  Length 2793;
  Best Local Similarity   50.9%;  Pred. No. 2.3e-15;
  Matches  269;  Conservative    0;  Mismatches  260;  Indels    0;  Gaps    0;

Qy   1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553

Db   2120 CAGCGGCTCCAGCTCCTGCAGCGGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGC 2061

Qy   1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613

Db   2060 CTCCGGATCCTGCACCAGCAGCTGCAGCAGCACCAGCTCCTGCTCCGGCACCGTATCCAC 2001

Qy   1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673

Db   2000 CTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCAC 1941

Qy   1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733

Db   1940 CAGCGGCTCCTGCACGACCACCTCTACTGTAACCCGCAGCGCCTCCAGCTCCTGCTCCAG 1881

Qy   1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793

Db   1880 CAGCTGCCGCAGCACCAGCTCTTGCTCCGGCACCGTATCCACCCTGACCACCGTAGCCTC 1821

Qy   1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853

Db   1820 CAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCACCTGCAGCAGCTCCAGCTC 1761

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913

Db   1760 CAGCACCAGCACCTCTACCGTAACCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCG 1701

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973

Db   1700 CACCAGCTCCTGCTCCGGCACCATAACCGCCTTGTCCACCATATCCAGCACCACCGGTGC 1641

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCC 2022

Db   1640 CTGCACCTGCCCCAGCAGCGGCACCAGCTCCAGCACCAGCACCTCTACC 1592


RESULT   9
US-08-770-379-20
; Sequence 20, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  32207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-20


  Query Match             4.4%;  Score 108.6;  DB 4;  Length 32207;
  Best Local Similarity   48.5%;  Pred. No. 3.7e-14;
  Matches  300;  Conservative    0;  Mismatches  319;  Indels    0;  Gaps    0;

Qy    1503 CTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTG 1562
           || | |  ||  |||||||  | ||| | ||  |  |||| |   |  || | ||  ||
Db   19348 CTCCTCTAACTCCTGCTCCTCCTGCTCCTCCTGCTCCTGCTCTTGCTCCTCCACCTCCTC 19407

Qy    1563 CCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCAC 1622
             ||  |       |  ||  ||    | |  ||  |  ||||     |  | ||||
Db   19408 TAATTCCTGCTCTTCCTGCTCCTGCTCTTGCTCTTCCACCTCCTGCTCTTGCTCTTCCAC 19467

Qy    1623 CTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTG 1682
           || |  |  |  |  || || |  |  |  |      | ||  |  | ||||||   |||
Db   19468 CTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTCTAACTC 19527

Qy    1683 CTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCC 1742
           ||||  ||||   | | |||  |||   | |  || |  ||| |||||||  |  |
Db   19528 CTGCTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTCTAA 19587

Qy    1743 CTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCA 1802
           || || || |||  |||  ||  ||     |||   |   | | ||| |  ||||  ||
Db   19588 CTCCTGCTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTCTAACTCCTGCTCCTGCTCCTC 19647

Qy    1803 CCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAA 1862
             || |  | | | | | |  ||  | || ||| || ||      || |      |
Db   19648 TAACTCCTGCTCCTGATCCTCTAACTCCTGCTCCTGCTCCTCTAACTCCTGCTCCTGCTC 19707

Qy    1863 CTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCA 1922
           || |||   |  |  | ||| |||||||  ||   |    | ||  ||||  |  |  |
Db   19708 CTCCTGCTGCTCCTGCTCCTCCTGCTGCTCCTGTTCATCCTGCTGCTGCTGCTCATCCTG 19767

Qy    1923 CAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAG 1982
           | |   |||  |  ||  |||||||  ||  | | || ||  |||| |       |  | |
Db   19768 CTGCTGCTGCTCATCCTGCTGCTGCTGCTCATCCTGCTGCTGCTGCTCATCCTGCTGCTG 19827

Qy    1983 TCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCA 2042
             ||  ||  |  |||||   ||  ||||  || | | | | |  || | ||| |
Db   19828 CTGCTCATCCTGCTGCTGCTCATCCTGCTGCTCCTGCTCATCCTGCTGCTCCTGCTCATC 19887

Qy    2043 ACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATG 2102
             |  |  |||| |      || | |     |  | ||| |    ||||  | || |
Db   19888 CTGCTGCTCCTGCTCATCCTGCTGCTGCTCATCCTGCTGCTGCTCATCCTGCTGCTGCTC 19947

Qy    2103 ACATCGACCCCCCTCTCCC 2121
           |    |   |  |||  ||
Db   19948 ATCCTGCTGCTGCTCATCC 19966


RESULT  10
US-08-056-200-93/c
; Sequence 93, Application US/08056200
; Patent No. 5616500
;  GENERAL INFORMATION:
;    APPLICANT:  Steinert, Peter M.
;    APPLICANT:  Lee, Seung-Chul
;    APPLICANT:  Kim, In-Gyu
;    APPLICANT:  Chung, Soo-Il
;    APPLICANT:  Park, Sang-Chul
;    TITLE OF INVENTION:  Trichohyalin and Transglutaminase-3 and
;    TITLE OF INVENTION:  Mehods of Using Same
;    NUMBER OF SEQUENCES:  117
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/056,200
;      FILING DATE:  30-APR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fedrick, Michael F.
;      REGISTRATION NUMBER:  36,799
;      REFERENCE/DOCKET NUMBER:  NIH054.001A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (714) 760-0404
;      TELEFAX:  (714) 760-9502
;  INFORMATION FOR SEQ ID NO:  93:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9551 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1507..1644
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1645..2511
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2512..8070
US-08-056-200-93


  Query Match             4.1%;  Score 100.6;  DB 1;  Length 9551;
  Best Local Similarity   48.7%;  Pred. No. 1.2e-12;
  Matches  274;  Conservative    0;  Mismatches  289;  Indels    0;  Gaps    0;

Qy    1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553
           || |  |||||  | |  |  ||||||   |  |||| |   |  || | ||  | |||
Db    3771 CAGCCGCTGCTCGCGCCTCTCCTGCTCGTGCTTCTGCTCGTGCCTCTCCTCCTCCTGCTC 3712

Qy    1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613
             | |  | || ||  |   | ||   |   |  ||     |  || | ||  ||   |||
Db    3711 GCGCCTCAGCTGCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCAGCTGCTGCTCGCGCCT 3652

Qy    1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673
           |||   |  ||    || ||  | || |||  ||| |  | |||||  | |  || |||||
Db    3651 CAGCTGCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCAGCTG 3592

Qy    1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733
           || ||   ||  |  |||   | |   || | |   | |  |   |   | | |||||| |
Db    3591 CTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCAGCTGCTGCTC 3532

Qy    1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793
             |||  | ||| || |||||  | |   |       |     | | |  |  | ||| |
Db    3531 GCGCCTCTCCTCCTCCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCTCCTCCTCCTGCTC 3472

Qy    1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
             ||| |  | |||||  | | || || | ||  | ||   | |  |  ||||||  |  |
Db    3471 GCGCCTCAGCTGCTGCTCGCGCCTCTCCTCCTCCTGCTCGCGCCTCAGCTGCTGCTCGCG 3412

Qy    1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
               ||  || ||  | | ||| |  || ||   ||| |  || | || || |  |||
Db    3411 CCTCTCCTCCTGCTCGCGCCTCAGCTGCTGCTCGCGCCTCTCCTCCTGCTGCTCGCGCCT 3352
```

```
Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
Db   3351 CTCCTGCTGCTCGCGCCTCTCCTCCTGCTGCTCGCGCCTCTCCTCCTCCTGCTTGCGCCT 3292

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
Db   3291 TAGTTGCTGCTCGCGCCTCAGCCTTTGCTGCTGCTGCTCTTCCTCCTGGCGCTCCCTCCT 3232

Qy   2034 ATGCACCCAACACCACAGCTGCC 2056
Db   3231 CAGCTCTTGCCGCTCCAGCTTCC 3209
```

RESULT 11
PCT-US93-03993-1/c
; Sequence 1, Application PC/TUS9303993
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; TITLE OF INVENTION: CONSTITUTIVE AND INDUCIBLE EPIDERMAL
; TITLE OF INVENTION: VECTOR SYSTEMS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03993
; FILING DATE: 19930428
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5405
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6530 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US93-03993-1

Query Match 4.0%; Score 99.4; DB 5; Length 6530;
Best Local Similarity 48.5%; Pred. No. 2e-12;
Matches 308; Conservative 0; Mismatches 321; Indels 6; Gaps 1;

```
Qy   1497 CGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCC 1556
Db   3595 CCACCGCAGCTGGAGCCTCCGCCCCCGGAGTAGCCGCCGCCGCAGCTGGAGCCACCGCTA 3536

Qy   1557 CAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAG 1616
Db   3535 GAACCGCCTCCGTAGCTCTGGCACTGATACTGTTGGGACGATCCGCCGCCGCCGCCACCG 3476

Qy   1617 TCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTA 1676
```

```
Db   3475 GAGGAACCACCTCCGCAGCTCGTGCCGCCGCCACCGGAATAGCCGCCGCCGCCGCCGCAG 3416

Qy   1677 CTACTGCTGCCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCG 1736
Db   3415 CTGGATCCACCGCCACCTGAGTAGCCGCCGCCACAGCTAGAGCCGCCGCTGCTTCCACCG 3356

Qy   1737 CAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTG 1796
Db   3355 CCGCCGGAATAGCCGCCTCCGCAGCTGGAGCCACCGCCGCCACCAGAGTACT------TG 3302

Qy   1797 CCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAA 1856
Db   3301 CCGCCTCCGGAACCGCCGCCGCAGCCACCGGAGCTGCCGCCTCCAGAGGAGCCTCCGCAG 3242

Qy   1857 TCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCA 1916
Db   3241 TAGGAGCCGCCGCCTCCTGATCCGCCTCCGCAGCTGGAGCCGCCGCCGCCGCCCCCGGAG 3182

Qy   1917 ACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACA 1976
Db   3181 GAGCCGCCGCCGCAGCTAGAGCCGCCGCCGCCACCACCTCCGGAGTACTTGACGCCCCCA 3122

Qy   1977 CCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATG 2036
Db   3121 CCAGAGCCGCCACCGCCTCCACAGCTGGAGCCACCCCCGGAGTAGCTGCCGCCGCCGCAG 3062

Qy   2037 CACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAG 2096
Db   3061 CTGGAGCCGCCGCCGCCTCCGGAGTAGCCACCTCCGCAGCTGGAGCCGCCGCCGCCCCCA 3002

Qy   2097 ATGATGACATCGACCCCCCTCTCCCCAACGACCCC 2131
Db   3001 GAGTAGCCACCGCCGCAGCTAGAGCCACCACCGCC 2967
```

RESULT 12
US-07-885-970A-25
; Sequence 25, Application US/07885970A
; Patent No. 5495070
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,970A
; FILING DATE: 19920518
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:

```
;      LENGTH:  1984 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium barbadense
;      STRAIN:  Sea Island
;    IMMEDIATE SOURCE:
;      LIBRARY:  EMBL SI
;      CLONE:  SIH6
US-07-885-970A-25


 Query Match             3.8%;  Score 93;  DB 1;  Length 1984;
  Best Local Similarity   53.1%;  Pred. No. 3.2e-11;
  Matches  198;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy   1593 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1652
          |||||||| || |  | ||  |    || ||| | ||  |      |  | ||  | || |
Db    424 CGCCACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 483

Qy   1653 CCGCTGCCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1712
          ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    484 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 543

Qy   1713 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1772
             | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||
Db    544 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 603

Qy   1773 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1832
          |||  |  | | ||| |  || | ||  | ||| | ||  |  ||||| |  |  ||
Db    604 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 663

Qy   1833 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1892
          |  || || |  ||  | |      ||| || |  | || | ||| | ||| || | |
Db    664 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 723

Qy   1893 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1952
          |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    724 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 783

Qy   1953 CTACCGCAGCACC 1965
          |      | ||||
Db    784 CAGTACAAACACC 796


RESULT  13
US-08-298-687A-25
; Sequence 25, Application US/08298687A
; Patent No. 5521078
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    TITLE OF INVENTION:  PLANTS FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/298,687A
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1985 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium barbadense
;      STRAIN:  Sea Island
;    IMMEDIATE SOURCE:
;      LIBRARY:  EMBL-SI
;      CLONE:  SIH6
US-08-298-687A-25


  Query Match             3.8%;  Score 93;  DB 1;  Length 1985;
  Best Local Similarity   53.1%;  Pred. No. 3.2e-11;
  Matches  198;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy   1593 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1652
          |||||||| || |  | ||  |    || ||| | ||  |      |  | ||  | || |
Db    424 CGCCACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 483

Qy   1653 CCGCTGCCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1712
          ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    484 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 543

Qy   1713 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1772
             | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||
Db    544 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 603

Qy   1773 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1832
          |||  |  | | ||| |  || | ||  | ||| | ||  |  ||||| |  |  ||
Db    604 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 663

Qy   1833 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1892
          |  || || |  ||  | |      ||| || |  | || | ||| | ||| || | |
Db    664 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 723

Qy   1893 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1952
          |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    724 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 783

Qy   1953 CTACCGCAGCACC 1965
          |      | ||||
Db    784 CAGTACAAACACC 796


RESULT  14
US-08-298-829-25
; Sequence 25, Application US/08298829
; Patent No. 5620882
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E.
;    TITLE OF INVENTION:  GENETICALLY ENGINEERING COTTON
;    TITLE OF INVENTION:  PLANTS FOR ALTERED FIBER
;    NUMBER OF SEQUENCES:  33
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;      STREET:  P.O. Box 2113, First Wisconsin Plaza
;      CITY:  Madison
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/298,829
;      FILING DATE:  19-OCT-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/885,970
;      FILING DATE:  18-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/617,239
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/253,243
;      FILING DATE:  04-OCT-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27,386
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (608) 283-2478
;      TELEFAX:  (608) 251-5139
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1985 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium barbadense
;      STRAIN:  Sea Island
;    IMMEDIATE SOURCE:
;      LIBRARY:  EMBL-SI
;      CLONE:  SIH6
US-08-298-829-25


  Query Match             3.8%;  Score 93;  DB 1;  Length 1985;
  Best Local Similarity   53.1%;  Pred. No. 3.2e-11;
  Matches  198;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy   1593 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1652
          ||||||||  ||  |  | ||  |    || ||| | ||  |     |  | ||  | || |
Db    424 CGCCACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 483

Qy   1653 CCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1712
          ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    484 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 543

Qy   1713 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1772
             | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||
Db    544 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 603

Qy   1773 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1832
          |||   |  |  | ||| |  || | ||  | ||| | ||  |  |||||  |  |  ||
Db    604 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 663

Qy   1833 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1892
          |  || || |  ||  | |       ||| || |   | || | ||| | ||| || | |
Db    664 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 723

Qy   1893 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1952
          |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    724 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 783

Qy   1953 CTACCGCAGCACC 1965
          |      | ||||
Db    784 CAGTACAAACACC 796


RESULT  15
US-08-217-327-3
; Sequence 3, Application US/08217327
; Patent No. 5474925
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E
;    APPLICANT:  Barton, Kenneth A
;    TITLE OF INVENTION:  Immobilized Proteins in Cotton Fiber
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O. Box 2113
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  USA
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/217,327
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/812,233
;      FILING DATE:  19-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990831
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Gossypium hirsutum
;      STRAIN:  Coker 312
;      TISSUE TYPE:  Fiber cells
;    IMMEDIATE SOURCE:
;      CLONE:  H6
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  72..716
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  72..147
US-08-217-327-3


  Query Match             3.7%;  Score 91.4;  DB 1;  Length 913;
  Best Local Similarity   52.8%;  Pred. No. 5.5e-11;
  Matches  197;  Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy   1593 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1652
```

```
Db    175 CGACACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 234

Qy   1653 CCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1712

Db    235 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 294

Qy   1713 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1772

Db    295 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 354

Qy   1773 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1832

Db    355 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 414

Qy   1833 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1892

Db    415 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 474

Qy   1893 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1952

Db    475 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 534

Qy   1953 CTACCGCAGCACC 1965

Db    535 CAGTACAAACACC 547


Search completed: July 18, 1999, 07:53:30
Job time: 311 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 10:32:10 ; Search time 133.54 Seconds
(without alignments)
3458.605 Million cell updates/sec

Title: US-09-294-663-1
Perfect score: 2455
Sequence: 1 GTAACGTTAAGTGAAAAGAA.....AAAAAAAAAAAAAAAAAAAA 2455

Scoring table: IDENTITY_NUC

Searched: 240622 seqs, 94065609 residues

Database : N_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	153.6	6.3	799	1	V55831	Nucleotide sequenc
c 2	151.6	6.2	10596	1	Q51731	Plasmid pCisEBON f
c 3	151.6	6.2	10596	1	T40348	Plasmid pCisEBON f
c 4	151.6	6.2	9600	1	V21683	Vector plasmid pCM
c 5	142.4	5.8	795	1	V55830	FLGA insert stabil
6	125.8	5.1	5163	1	V20700	Cryptosporidium pa
7	124.2	5.1	5318	1	V20701	Cryptosporidium pa
c 8	113	4.6	2744	1	Q98470	MiSP1-containing p
c 9	107.6	4.4	40875	1	T80043	Insert from cosmid
c 10	99.4	4.0	6530	1	Q51557	Loricrin gene. Con
11	98.8	4.0	502	1	Q06483	Intestinal mucin c
12	95.6	3.9	1368	1	N91235	DNA sequence of pr
13	93	3.8	860	1	Q06486	Intestinal mucin c
14	93	3.8	1984	1	T13030	Cotton fibre-speci
15	93	3.8	1985	1	T30250	Cotton fibre clone
16	93	3.8	1985	1	T70036	Cotton H6 gene and
17	91.4	3.7	859	1	Q12220	SMUC-53 intestinal
18	91.4	3.7	913	1	T07199	Cotton fiber-speci
19	91.4	3.7	913	1	T13034	Cotton fibre-speci
20	91.4	3.7	913	1	T30253	Cotton fibre cell-
21	91.4	3.7	913	1	T62610	Cotton fibre speci
22	91.4	3.7	913	1	T70041	Cotton fibre speci
23	90.8	3.7	502	1	Q12217	SMUC-40 intestinal
24	90	3.7	5008	1	Q27223	MSF precursor. New
c 25	88.2	3.6	3331	1	V22683	New DNA sequence i
c 26	88.2	3.6	2214	1	V22682	New DNA sequence i
27	86.6	3.5	836	1	Q06484	Intestinal mucin c
28	86.6	3.5	836	1	Q12218	SMUC-41 intestinal
29	86.2	3.5	1107	1	V38816	DNA encoding a dur
c 30	85.6	3.5	1505	1	Q55750	Genomic clone G11F
c 31	84.6	3.4	2000	1	N71065	Gene encoding Plas
c 32	83.4	3.4	44377	1	T78508	Platenolide syntha
c 33	83.4	3.4	13987	1	T80415	Hybrid srmG/tylG O
c 34	83.4	3.4	44377	1	T80414	Platenolide syntha
35	83	3.4	390	1	Q21833	Randomising oligon
36	83	3.4	390	1	Q36859	PCR primer for 5'
37	82.2	3.3	12001	1	Q76213	HSV L/ST region. H
38	78.4	3.2	2550	1	V52497	Streptococcus pneu
39	78	3.2	2367	1	V44448	Mycobacterium tube
40	78	3.2	2367	1	V64557	M. tuberculosis im
c 41	77.6	3.2	2338	1	Q14183	N.clavipes draglin
c 42	77.6	3.2	2338	1	V23249	Nephila clavipes s
c 43	75.8	3.1	2004	1	T85356	Nephila clavipes s
44	73.8	3.0	14704	1	Q20685	PKS 741 insert con
45	73.4	3.0	359	1	Q06487	Intestinal mucin c

ALIGNMENTS

RESULT 1
V55831/c
ID V55831 standard; DNA; 799 BP.
AC V55831;
DT 18-NOV-1998 (first entry)
DE Nucleotide sequence of the stabilising sequence-encoding insert.
KW Fusion protein; stabilising polypeptide; proteolytic degradation;
KW resistance; half-life; autoimmune disease; inflammation; nitro drug;
KW IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
KW nitroreductase protein; enzyme therapy; prodrug therapy; protease;
KW cancer; pathological condition; ss.
OS Epstein-barr virus.
PN WO9822577-A1.
PD 28-MAY-1998.
PF 17-NOV-1997; IB1508.
PR 25-JUN-1997; US-048945.
PR 15-NOV-1996; US-030986.
PA (MASU/) MASUCCI M G.
PI Masucci MG;
DR WPI; 98-312463/27.
PT New fusion proteins resistant to proteolytic degradation -
PT comprising a core protein with a stabilising polypeptide comprising
PT a peptide sequence containing glycine repeats
PS Disclosure; Fig 4B; 120pp; English.
CC This is a nucleotide sequence of the stabilising sequence-encoding
CC insert. The invention provides a method for increasing the resistance
CC of a core protein to proteolytic degradation that comprises linking or
CC inserting onto or into the core protein a stabilising polypeptide of
CC formula [(Glya)X(Glyb)Y(Glyc)Z]n where Glya, Glyb, Glyc are 1-6
CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC be identical from n repeat to n repeat. Alternatively a nucleic acid
CC encoding the stabilising polypeptide can be linked onto or inserted into
CC a nucleic acid encoding a core protein. The fusion proteins of the
CC invention are more resistant to degradation by proteases and, thus, have
CC a longer half-life than the unfused core protein. The products can be
CC used for treating autoimmune diseases, cancer and inflammation. In
CC particular, the core protein may be an IkappaB regulator protein for the
CC treatment of inflammatory bowel disease, or a nitroreductase protein
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC or other pathological conditions. The fusion proteins can also be used in
CC diagnostic methods such as in vivo imaging.
SQ Sequence 799 BP; 201 A; 106 C; 479 G; 13 T;

Query Match 6.3%; Score 153.6; DB 1; Length 799;
Best Local Similarity 53.2%; Pred. No. 1.7e-21;
Matches 352; Conservative 0; Mismatches 304; Indels 6; Gaps 1;

```
Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || |  |  |||||||  | ||| |  |  |  |||||||
Db    662 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 603

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  ||  |     | ||| ||| |  | ||  || |  |||| | || |  |
Db    602 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 543

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db    542 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 483

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||    |     | ||| |||| |||| || |
Db    482 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 423

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
```

```
Db    422 CTCCTGCTCCTGCCCCCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCTCCTG 363

Qy   1796 --GCCCCCACCGCTGCCCCCCACCGCTGCCCCCACCGCAGCCCCCTACTACTGCTGTCCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db    362 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCCTCCTGCCCCCTCCTCCTGCTCCTGCCCCCTC 303

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
             |||  || |||   |  | || ||| ||||| |  || || | ||  | |||||||
Db    302 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 243

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCCCCCA 1973
           |     |  |     |    ||  || || || |  || || || |  | ||||| ||
Db    242 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 183

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
            | |  | |.| ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db    182 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 123

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
            ||| ||     || |  |||||||     || |  |   | | ||  || |  |  |
Db    122 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 63

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAAT 2153
            || |     |   ||| |||   |    |  ||  |   .|| || || |  | ||
Db     62 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGTCGGGGGAT 3

Qy   2154 GC 2155
            |
Db      2 CC 1
```

```
RESULT   2
Q51731/c
ID   Q51731 standard; DNA; 10596 BP.
AC   Q51731;
DT   31-MAY-1994  (first entry)
DE   Plasmid pCisEBON for subcloning huHGF variants.
KW   Hepatocyte Growth Factor; HGF; variant; mutein; in vitro mutagenesis;
KW   proteolysis resistant; liver; malignancy; CMV-driven;
KW   Cytomegalovirus; episomal expression plasmid; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   enhancer        1. .611
FT                   /*tag= a
FT                   /note= "CMV enhancer/promoter"
FT   promoter        758. .775
FT                   /*tag= b
FT                   /label= SP6_promoter
FT   misc_feature    845. .849
FT                   /*tag= c
FT                   /note= "SP6 RNA start"
FT   misc_feature    902. .966
FT                   /*tag= d
FT                   /function= cloning_linker
FT   polya_signal    967. .1107
FT                   /*tag= e
FT                   /note= "SV40 poly A"
FT   misc_feature    1108. .1531
FT                   /*tag= f
FT                   /function= SV40_origin
FT   misc_feature    1580. .4189
FT                   /*tag= g
FT                   /label= EBNA-1
FT   misc_feature    4190. .6374
FT                   /*tag= h
FT                   /function= oriP
FT   repeat_region   4295. .4887
FT                   /*tag= i
FT                   /note= "family of repeats"
FT   misc_structure  5866. .5978
FT                   /*tag= j
FT                   /note= "dyad region"
FT   terminator      6375. .6457
FT                   /*tag= k
FT                   /label= HSV_TK_terminator_3'-end
FT   cds             6975. .7975
FT                   /*tag= l
FT                   /phenotype= neomycin_resistance
FT                   /note= "Tn5 neomycin phosphotransferase gene"
FT   promoter        7975. .8112
FT                   /*tag= m
FT                   /label= TK_promoter
FT   misc_feature    8114. .8594
FT                   /*tag= n
FT                   /function= M13_ori
FT   misc_rna        8595. .10414
FT                   /*tag= o
FT                   /label= delta_2a
PN   WO9323541-A.
PD   25-NOV-1993.
PF   17-MAY-1993; U04648.
PR   18-MAY-1992; US-884811.
PR   18-MAY-1992; US-885971.
PA   (GETH ) GENENTECH INC.
PI   Godowski PJ,  Lokker NA,  Mark MR;
DR   WPI; 93-386573/48.
PT   Hepatocyte growth factor variants - are resistant to proteolytic
PT   cleavage into its two-chain form, used to treat malignancies
PT   associated with HGF receptor
PS   Example 1; Fig 6; 87pp; English.
CC   Plasmid pCisEBON (a pRK5 derivative) is an episomal CMV driven
CC   expression plasmid. HuHGF variants with enhanced receptor binding
CC   activity were produced by site-directed mutagenesis. Stable
CC   populations of preferred HGF variants were obtained by transfecting
CC   human embryonic kidney 293 cells and then these were subcloned in
CC   pCisEBON. See R52940-R52949 for examples of pref. HGF variants.
SQ   Sequence   10596 BP;    2625 A;    2571 C;    3024 G;    2376 T;
```

Query Match 6.2%; Score 151.6; DB 1; Length 10596;
Best Local Similarity 53.2%; Pred. No. 7e-21;
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

```
Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
            || |||||||  | ||| || |  |  |||||||  | ||| |  |  |  |||||||
Db    2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
            || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db    2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
            |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db    2747 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 2688

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
            || ||||||||    | ||  || ||     |      | ||| |||| ||||  || |
Db    2687 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
            | ||| || ||||    ||||       |     ||| | |  | || ||| | |||
Db    2627 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCCTACTACTGCTGTCCCAG 1853
             | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db    2567 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
             |||  || |||   |  | || ||| ||||| |  || || | ||  | |||||||
Db    2507 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
            |     |  |     |    ||  || || || |  || || || |  | ||||| ||
Db    2447 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 2388
```

```
Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
          | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db   2387 CCCCTCCTGCCCCTCCTGCCCCCTCCTCCTGCTCCTGCCCCCTCCTCCTGCTCCTGCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
          |||  ||     ||  |  |||||||     ||  |  |   |  | ||  ||  |  |  |
Db   2327 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
               ||  |     |   |||  |||   |     |  ||  |    || ||  ||
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216


RESULT   3
T40348/c
ID   T40348 standard; DNA; 10596 BP.
AC   T40348;
DT   09-DEC-1996  (first entry)
DE   Plasmid pCisEBON for expression of hepatocyte growth factor.
KW   Human; hepatocyte growth factor; HGF; huHGF; serum; proteolytic cleavage;
KW   pro-hormone; beta subunit; alpha subunit; kringle domain; prothrombin;
KW   plasminogen; catalytic domain; serine protease; HGF variant;
KW   HGF receptor; malignancy; chronic HGF receptor activation; ss.
OS   Synthetic.
PN   US5547856-A.
PD   20-AUG-1996.
PF   18-MAY-1992; 884811.
PR   18-MAY-1992; US-885971.
PR   18-MAY-1992; US-884811.
PR   13-JUL-1993; US-087783.
PA   (GETH ) GENENTECH INC.
PI   Godowski PJ,  Lokker NA,  Mark MR;
DR   WPI; 96-392634/39.
PT   New hepatocyte growth factor variants - are resistant to in vivo
PT   proteolytic cleavage into a 2-chain form, useful as HGF antagonists
PS   Example 1; Fig 6; 39pp; English.
CC   This sequence represents the episomal CMV driven expression plasmid
CC   pCisEBON which was used in the expression of variant human hepatocyte
CC   growth factor (HGF).  HGF is isolated from human serum and is a
CC   disulphide linked heterodimer derived by proteolytic cleavage of the
CC   pro-hormone between residues 494 and 495.  This generates a molecule
CC   composed of an alpha subunit of 440 amino acids (mol. wt. 69 kD) and
CC   a beta subunit of 234 amino acids (mol. wt. 34 kD).  The alpha and beta
CC   subunits are encoded by a single open reading frame.  The alpha subunit
CC   contains four kringle domains based on their homology to kringle-like
CC   domains in other proteins, e.g. prothrombin, plasminogen.  The beta
CC   subunit shows high homology to the catalytic domain of serine proteases.
CC   However two of the three residues which form the catalytic triad of
CC   serine proteases are not conserved in HGF.  Therefore, the precise
CC   function of the beta chain remains unknown.  The invention includes HGF
CC   variants which retain HGF receptor binding activity without having the
CC   biological activity of wild-type HGF.  They can be used for the treatment
CC   of pathological conditions associated with the activation of a HGF
CC   receptor such as malignancies associated with chronic HGF receptor
CC   activation.  The pCisEBON plasmid comprising the HGF coding sequence
CC   may be used for manipulation of the HGF coding sequence and expression
CC   of the variant HGF's of the invention.
SQ   Sequence   10596 BP;    2627 A;    2571 C;    3023 G;    2375 T;


  Query Match             6.2%;  Score 151.6;  DB 1;  Length 10596;
  Best Local Similarity   53.2%;  Pred. No. 7e-21;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          ||  |||||||  | |||  || |  |  |||||||  | |||  |  |  |  |||||||
Db   2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          ||  |  ||  ||  ||  |    | |||  |||  |  |  ||  ||  |  ||||  | ||  |  |
Db   2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   ||||  ||||  |||||  | ||  | ||  |  | || ||| ||
Db   2747 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 2688

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          ||  ||||||||     | ||  || ||      |      | |||  ||||  ||||  || |
Db   2687 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT----, 1796
          | |||  ||  ||||     ||||        |       |||  | |   |  ||  |||  | |||
Db   2627 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            |  ||   | |||  |   ||  |   |||||||   |  ||   | |||  ||  ||  |||  |||
Db   2567 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
               |||  ||  |||    |   |  ||  |||  |||||  |   ||  ||  |  ||   | ||||||||
Db   2507 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |     |  |     |    ||  ||  ||  ||  |   ||  ||  ||  |   |  |||||  ||
Db   2447 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 2388

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
          | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db   2387 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
          |||  ||     ||  |  |||||||     ||  |  |   |  | ||  ||  |  |  |
Db   2327 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
               ||  |     |   |||  |||   |     |  ||  |    || ||  ||
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216


RESULT   4
V21683/c
ID   V21683 standard; DNA; 9600 BP.
AC   V21683;
DT   17-AUG-1998  (first entry)
DE   Vector plasmid pCMVKmITR-EPI.
KW   Polynucleotide delivery; plasmid pCMVKmITR-EPI; vector;
KW   gene therapy; vaccine; polycationic agent; ss.
OS   Chimeric - Epstein-Barr virus.
OS   Chimeric - Adeno-associated virus.
OS   Chimeric - Cytomegalovirus.
OS   Chimeric - Bos taurus.
FH   Key             Location/Qualifiers
FT   CDS             14. .2594
FT                   /*tag=  a
FT                   /product= "EBV nuclear antigen A"
FT   misc_feature    2623. .4559
FT                   /*tag=  b
FT                   /note= "EBV origin of replication"
FT   repeat_unit     4928. .5104
FT                   /*tag=  c
FT                   /rpt_type= INVERTED
FT                   /note= "AAV inverted terminal repeat"
FT   repeat_unit     7189. .7355
FT                   /*tag=  d
FT                   /rpt_type= INVERTED
FT                   /note= "AAV inverted terminal repeat"
FT   promoter        5112. .6734
FT                   /*tag=  e
FT                   /note= "CMV immediate-early enhancer/promoter"
FT   terminator      6818. .7050
FT                   /*tag=  f
FT                   /note= "bovine growth hormone polyA sequence"
PN   WO9806437-A2.
PD   19-FEB-1998.
PF   13-AUG-1997; U14465.
PR   13-AUG-1996; US-023867.
```

PA (CHIR) CHIRON CORP.
PI Cohen F, Dubois-Stringfellow N, Dwarki V, Innis MA,
PI Murphy JE, Tetsuo U, Zukermann R;
DR WPI; 98-159296/14.
PT Polycationic agents based on alpha-amino acids, able to complex
PT with nucleic acid - to facilitate its entry into cell, condense it
PT and protect it against serum degradation, particularly for use in
PT gene therapy
PS Disclosure; Page 77-80; 100pp; English.
CC This polynucleotide comprises the DNA sequence of vector plasmid
CC pCMVKmITR-EPI, which contains an Epstein-Barr virus (EBV) origin
CC of replication from plasmid pCEP4, a coding region for EBV nuclear
CC antigen A from pCEP4, a pair of inverted terminal repeats from
CC adeno-associated virus, a cytomegalovirus enhancer/promoter, a
CC bovine growth hormone polyA sequence, and a kanamycin resistance
CC selectable marker. Polynucleotides encoding polypeptides, such as
CC erythropoietin or leptin, and ribozymes and antisense
CC polynucleotides can be inserted into the vector. The vector is
CC preferred for use in novel compositions and methods for improved
CC polynucleotide delivery into cells. In these methods, polycationic
CC agents are used to increase the frequency of uptake of a
CC nucleic acid (see also V21684-86) into a cell. The polycationic
CC agent can condense with the nucleic acid and inhibit serum and/or
CC nuclease degradation of the nucleic acid. The nucleic acid can be
CC a vector, may express a therapeutic protein or a vaccinating viral
CC or cancer antigen, or is itself therapeutic (antisense or
CC ribozyme). The methods and compositions can be used in the gene
CC therapy of many diseases.
SQ Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T;

Query Match 6.2%; Score 151.6; DB 1; Length 9600;
Best Local Similarity 53.2%; Pred. No. 6.8e-21;
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

Qy 1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
Db 1333 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 1274

Qy 1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
Db 1273 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 1214

Qy 1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
Db 1213 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 1154

Qy 1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
Db 1153 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 1094

Qy 1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
Db 1093 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 1034

Qy 1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
Db 1033 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 974

Qy 1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
Db 973 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 914

Qy 1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
Db 913 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 854

Qy 1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
Db 853 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 794

Qy 2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
Db 793 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 734

Qy 2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
Db 733 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 682

RESULT 5
V55830/c
ID V55830 standard; DNA; 795 BP.
AC V55830;
DT 18-NOV-1998 (first entry)
DE FLGA insert stabilising polypeptide encoding DNA.
KW Fusion protein; stabilising polypeptide; proteolytic degradation;
KW resistance; half-life; autoimmune disease; inflammation; nitro drug;
KW IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
KW nitroreductase protein; enzyme therapy; prodrug therapy; protease;
KW cancer; pathological condition; ss.
OS Epstein-barr virus.
FH Key Location/Qualifiers
FT CDS 1. .788
FT /*tag= a
FT /product= "stabilising polypeptide""
PN WO9822577-A1.
PD 28-MAY-1998.
PF 17-NOV-1997; IB1508.
PR 25-JUN-1997; US-048945.
PR 15-NOV-1996; US-030986.
PA (MASU/) MASUCCI M G.
PI Masucci MG;
DR WPI; 98-312463/27.
DR P-PSDB; W79128.
PT New fusion proteins resistant to proteolytic degradation -
PT comprising a core protein with a stabilising polypeptide comprising
PT a peptide sequence containing glycine repeats
PS Disclosure; Fig 3; 120pp; English.
CC This DNA encodes a stabilising polypeptide and is the FLGA insert of the
CC invention. The invention provides a method for increasing the resistance
CC of a core protein to proteolytic degradation that comprises linking or
CC inserting onto or into the core protein a stabilising polypeptide of
CC formula [(Glya)X(Glyb)Y(Glyc)Z]n where Glya, Glyb, Glyc are 1-6
CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC be identical from n repeat to n repeat. Alternatively a nucleic acid
CC encoding the stabilising polypeptide can be linked onto or inserted into
CC a nucleic acid encoding a core protein. The fusion proteins of the
CC invention are more resistant to degradation by proteases and, thus, have
CC a longer half-life than the unfused core protein. The products can be
CC used for treating autoimmune diseases, cancer and inflammation. In
CC particular, the core protein may be an IkappaB regulator protein for the
CC treatment of inflammatory bowel disease, or a nitroreductase protein
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC or other pathological conditions. The fusion proteins can also be used in
CC diagnostic methods such as in vivo imaging.
SQ Sequence 795 BP; 200 A; 104 C; 478 G; 13 T;

Query Match 5.8%; Score 142.4; DB 1; Length 795;
Best Local Similarity 53.2%; Pred. No. 2.5e-19;
Matches 351; Conservative 0; Mismatches 301; Indels 8; Gaps 2;

Qy 1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
Db 659 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 600

Qy 1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
Db 599 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 540

Qy 1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
Db 539 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 480

```
Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||     |     | ||| ||||  ||||  || |
Db    479 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 420

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||       |     ||| | |  | || ||| | |||
Db    419 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 360

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db    359 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 300

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
          |||  || |||   |  | || ||| |||||  |  || || || |  | |||||||
Db    299 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCCCC--CTCCTGCCC 242

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |     |  |     |    ||  || || || |  || || || |  | |||||  ||
Db    241 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 182

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
          | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db    181 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 122

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
           ||| ||     || |  |||||||     || |   |   | | ||  || |  |  |
Db    121 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 62

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAAT 2153
             || |     |   ||| |||   |    |  ||  |    || || || |  | ||
Db     61 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGTCGGGGGAT 2

RESULT   6
V20700
ID   V20700 standard; DNA; 5163 BP.
AC   V20700;
DT   17-AUG-1998  (first entry)
DE   Cryptosporidium parvum GP900 antigen open reading frame.
KW   Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW   antibody; prophylaxis; treatment; inhibition; retardation;
KW   detection; diagnosis; human; ds.
OS   Cryptosporidium parvum.
FH   Key             Location/Qualifiers
FT   CDS             1. .5163
FT                   /*tag=  a
FT                   /product= GP900 antigen
FT   misc_feature    524. .1270
FT                   /*tag=  b
FT                   /note= "region containing NINC mutations"
PN   WO9806430-A1.
PD   19-FEB-1998.
PF   11-AUG-1997; U14104.
PR   14-AUG-1996; US-700651.
PA   (REGC ) UNIV CALIFORNIA.
PI   Gut J, Leech J, Nelson RC, Petersen C;
DR   WPI; 98-159290/14.
DR   P-PSDB; W48299.
PT   Anti-Cryptosporidium antibody - used to develop products for
PT   detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT   infections
PS   Claim 32; Pages 60-62; 89pp; English.
CC   The sequence is that encoding the GP900 antigen which may be used
CC   in the production of anti-Cryptosporidium antibodies. These can be
CC   used for the prophylaxis, treatment, inhibition or retardation of
CC   a Cryptosporidium infection in humans or in animals such as calves.
CC   They can also be used for the detection and diagnosis of related
CC   infections.
SQ   Sequence   5163 BP;    1875 A;    1137 C;    873 G;    1278 T;

  Query Match             5.1%;  Score 125.8;  DB 1;  Length 5163;
  Best Local Similarity   51.6%;  Pred. No. 5.6e-16;

  Matches  344;  Conservative    0;  Mismatches  312;  Indels   11;  Gaps    2;

Qy   1493 ACATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCT 1552
          |||  | ||  || |  | ||  || || | ||  |  |  ||||  |  |  ||||  |
Db    584 ACAACAACAACTACTACTACTACTACTACGACAACAACAACAACAACGACAACAACAACA 643

Qy   1553 GCCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCC 1612
          |  | ||  |  |  || | ||        || |  ||||  || |  ||||  || |
Db    644 ACTACTACAACTACTACCACTACTACTACGACAACAACAACAACAACAACAACAACAACA 703

Qy   1613 CCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCT 1672
          ||    || | |||||  |||   | ||  || |  | ||  || |  | ||  |  ||
Db    704 ACAACAACAACAACAACAACAACAACAACAACGACTACTACTACTACTACTACTACTACT 763

Qy   1673 CCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCC 1732
          ||||||||  || |  |      |  |||| ||   ||||    || | ||  |  || |
Db    764 ACTACTACTACTACTACAACCACAACAACTACAACCACAACTACAACCACAACAACTACA 823

Qy   1733 ACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACC 1792
          |||  ||  |  |  || ||  |      ||  ||  | |||  ||   |||| || |  | |||
Db    824 ACCACAACAACAACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACC 883

Qy   1793 GCTGCCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCA 1852
          |  |  ||||  |  |  ||||  || ||   |   || |  | || ||| |
Db    884 ACAACAACCACAACCACAACCACAACTACCAAGAAACCAACAACAACAACTAC------- 937

Qy   1853 GAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCC 1912
            ||   |||| ||    ||| || |  ||||| |  || |||| ||  || || |  |
Db    937 --AACAACAACAACAACAACAACAACTACTACTACAACCACCACAACAACAACAACAACA 994

Qy   1913 CCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCC 1972
          | |  || ||   ||    |||| | ||| || |  |||||||| || || |  |  |
Db    995 ACAACTACAACTACCAAGAAACCAACAACTACTACTACTACTACCACAACAACAACAACT 1054

Qy   1973 AACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCA 2032
          |  || ||   |||   | || | ||| || |  | ||| || |  |  | ||     ||
Db   1055 ACTACTACTACCACAACAACAACAACTACTACTACTACTACAACAACAACAACAACAACA 1114

Qy   2033 CATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACC- 2092
            || | |  || ||  |    ||   ||| ||||| |   |  |  |||| ||  |
Db   1115 ACAACAACAACAACTACCACGAAACCAACAACAACAACAACAACTACTACTACTACTACA 1174

Qy   2092 -TGAAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAG 2150
           ||||     ||| | ||  ||    || |||| ||  |  | ||   ||      ||
Db   1175 ACCAAGAAACCAACAACAACTACCACTGCCACAACAACAACTACTACTTCTGAAACTGAG 1234

Qy   2151 AATGCAA 2157
          | || ||
Db   1235 AGTGTAA 1241

RESULT   7
V20701
ID   V20701 standard; DNA; 5318 BP.
AC   V20701;
DT   17-AUG-1998  (first entry)
DE   Cryptosporidium parvum GP900 antigen open reading frame and 3' region.
KW   Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW   antibody; prophylaxis; treatment; inhibition; retardation;
KW   detection; diagnosis; human; 3' region; ds.
OS   Cryptosporidium parvum.
FH   Key             Location/Qualifiers
FT   CDS             1. .5166
FT                   /*tag=  a
FT                   /product= GP900 antigen
FT   misc_feature    524. .1270
FT                   /*tag=  b
FT                   /note= "region containing NINC mutations"
PN   WO9806430-A1.
PD   19-FEB-1998.
PF   11-AUG-1997; U14104.
PR   14-AUG-1996; US-700651.
PA   (REGC ) UNIV CALIFORNIA.
```

```
PI   Gut J, Leech J, Nelson RC, Petersen C;
DR   WPI; 98-159290/14.
DR   P-PSDB; W48299.
PT   Anti-Cryptosporidium antibody - used to develop products for
PT   detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT   infections
PS   Claim 32; Pages 62-63; 89pp; English.
CC   The sequence is that encoding the GP900 antigen which may be used
CC   in the production of anti-Cryptosporidium antibodies. These can be
CC   used for the prophylaxis, treatment, inhibition or retardation of
CC   a Cryptosporidium infection in humans or in animals such as calves.
CC   They can also be used for the detection and diagnosis of related
CC   infections.
SQ   Sequence   5318 BP;    1944 A;    1153 C;    888 G;    1333 T;


  Query Match             5.1%;  Score 124.2;  DB 1;  Length 5318;
  Best Local Similarity   51.4%;  Pred. No. 1.1e-15;
  Matches  343;  Conservative    0;  Mismatches  313;  Indels   11;  Gaps    2;

Qy   1493 ACATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCT 1552
           ||| | ||  || | | || || || | ||  |  |  ||||  |  |  ||||  |
Db    584 ACAACAACAACTACTACTACTACTACTACGACAACAACAACAACGACAACAACAACA 643

Qy   1553 GCCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCC 1612
           |  | ||  |  |  || | ||          || |  ||||   | |  ||||  | |
Db    644 ACTACTACAACTACTACCACTACTACTACGACAACAACAACAAAAACAACAACAACAACA 703

Qy   1613 CCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCT 1672
           ||    || | |||||  |||   | ||  || |  | ||  || |  | ||  |  ||
Db    704 ACAACAACAACAACAACAACAACAACAACAACGACTACTACTACTACTACTACTACTACT 763

Qy   1673 CCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCC 1732
           |||||||| || |  |     | |||| ||   ||||    || | ||  |  | || |
Db    764 ACTACTACTACTACTACAACCACAACAACTACAACCACAACTACAACCACAACAACTACA 823

Qy   1733 ACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACC 1792
           |||  || |  | || ||  |    ||  || | |||  ||   |||| || |  | |||
Db    824 ACCACAACAACAACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACC 883

Qy   1793 GCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCA 1852
           |  |  ||||  |  |  ||||  || ||   |   || |  | || ||| |
Db    884 ACAACAACCACAACCACAACCACAACTACCAAGAAACCAACAACAACAACTAC------- 937

Qy   1853 GAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCC 1912
             ||   |||| ||    ||| || |  ||||| |  || |||| ||  || || |  |
Db    937 --AACAACAACAACAACAACAACTACTACTACAACCACCACAACAACAACAACAACA 994

Qy   1913 CCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCC 1972
           | |  || ||   ||    |||| | ||| || |  |||||||| || || |  |  |
Db    995 ACAACTACAACTACCAAGAAACCAACAACTACTACTACTACTACCACAACAACAACAACT 1054

Qy   1973 AACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCA 2032
           |  || ||   |||   | || | ||| || |  | ||| || | |  | || |
Db   1055 ACTACTACTACCACAACAACAACAACTACTACTACTACTACAACAACAACAACAACAACA 1114

Qy   2033 CATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACC- 2092
             || | |  || ||  |    ||  ||| ||||| |   |  |  |||| ||  |
Db   1115 ACAACAACAACAACTACCACGAAACCAACAACAACAACAACAACTACTACTACTACTACA 1174

Qy   2092 -TGAAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAG 2150
            ||||     ||| | ||  ||   || |||| ||  |  | ||   ||      ||
Db   1175 ACCAAGAAACCAACAACAACTACCACTGCCACAACAACAACTACTACTTCTGAAACTGAG 1234

Qy   2151 AATGCAA 2157
           | || ||
Db   1235 AGTGTAA 1241


RESULT   8
Q98470/c
ID   Q98470 standard; cDNA; 2744 BP.
AC   Q98470;
DT   01-MAY-1996  (first entry)
DE   MiSP1-containing plasmid pMISS1.
KW   Spider silk; repeat unit; consensus; minor ampullate silk protein;
KW   spidroin; MiSP; orb web spider; dragline; ds.
OS   Nephila clavipes.
FH   Key             Location/Qualifiers
FT   cds             183. .2678
FT                   /*tag= a
FT   misc_difference 996. .1037
FT                   /*tag= b
FT                   /note= "represented as indeterminate, a highly
FT                   compressed GC rich region which could
FT                   not be sequenced"
PN   WO9525165-A1.
PD   21-SEP-1995.
PF   14-MAR-1995; U03139.
PR   14-MAR-1994; US-209747.
PA   (UYWY-) UNIV WYOMING.
PI   Colgin M,  Lewis RV;
DR   WPI; 95-336970/43.
DR   P-PSDB; R80168.
PT   Polypeptide(s) comprising repeated unit amino acid sequences, also
PT   cDNAs - derived from minor ampullate spider silk proteins and used
PT   to form spider silk fibres
PS   Claim 12; Fig 1; 86pp; English.
CC   cDNA clone, pMISS1, encoding the orb web spider minor ampullate silk
CC   protein MiSP1 has been identified and sequenced. Repeat unit peptides
CC   of MiSP1 may form part of a larger polypeptide with an amino terminus
CC   (see R80184-85) and a carboxy terminus (see R80186-89). Repeat unit
CC   peptides make up spider silk proteins (spidroins) which in turn
CC   aggregate to form the silk fibres. Spider silk fibres have high
CC   tensile strength and significant elasticity. An isolated cDNA clone of
CC   a silk protein encoding sequence is of use to produce the protein at
CC   high yields using recombinant DNA technology.
SQ   Sequence   2744 BP;    486 A;    549 C;    1059 G;    608 T;


  Query Match             4.6%;  Score 113;  DB 1;  Length 2744;
  Best Local Similarity   50.9%;  Pred. No. 1.4e-13;
  Matches  269;  Conservative    0;  Mismatches  260;  Indels    0;  Gaps    0;

Qy   1494 CATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTG 1553
           || || || | || ||  |||| |||||| |  | || ||      |    ||| | ||
Db   2121 CAGCGGCTCCAGCTCCTGCAGCGGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGC 2062

Qy   1554 CCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCC 1613
           | ||   | | ||||  |  | |  |   |||| ||  |  | || ||| |    | |
Db   2061 CTCCGGATCCTGCACCAGCAGCTGCAGCAGCACCAGCTCCTGCTCCGGCACCGTATCCAC 2002

Qy   1614 CAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTC 1673
           |    |||||     |  ||  |||| |||| || ||  | || || ||  |  |||| |
Db   2001 CTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCAC 1942

Qy   1674 CTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA 1733
           |  |  || |||| |    | | | |         | |||  || ||  || |||||||
Db   1941 CAGCGGCTCCTGCACGACCACCTCTACTGTAACCCGCAGCGCCTCCAGCTCCTGCTCCAG 1882

Qy   1734 CCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCG 1793
           | || ||| |  |  | ||| |  ||       ||      |   |  | || ||    |
Db   1881 CAGCTGCCGCAGCACCAGCTCTTGCTCCGGCACCGTATCCACCCTGACCACCGTAGCCTC 1822

Qy   1794 CTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
           | || ||  |  |||| ||  | || || ||  |  |||| ||| |  | |||    |
Db   1821 CAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCAGCACCTGCAGCAGCTCCAGCTC 1762

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
             |  ||| |  || |  | | |||| |  || |||| ||  |  | ||||||  | |||
Db   1761 CAGCACCAGCACCTCTACCGTAACCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCG 1702

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
           |     |  | |  | | ||||    |  ||   ||  |     |||||||  |
Db   1701 CACCAGCTCCTGCTCCGGCACCATAACCGCCTTGTCCACCATATCCAGCACCACCGGTGC 1642
```

```
Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCC 2022
          |  | | | | | | |  |  | ||| |  |  | ||| |||  ||
Db   1641 CTGCACCTGCCCCAGCAGCGGCACCAGCTCCAGCACCAGCACCTCTACC 1593


RESULT   9
T80043/c
ID   T80043 standard; DNA; 40875 BP.
AC   T80043;
DT   03-NOV-1997  (first entry)
DE   Insert from cosmid 109.
KW   Pectinase; yeast; polygalacturonase; phytopathogen; vegetable spoilage;
KW   fruit spoilage; food processing; fruit juice; cosmid; ss.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   cds             312. .1943
FT                   /*tag=  a
FT                   /note= "indicated in the specification as an open reading
FT                   frame, encoding a protein with homology to
FT                   S. cerevisiae allantoate permease (DAL5)"
FT   cds             3622. .7107
FT                   /*tag=  b
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             3965. .5188
FT                   /*tag=  c
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             3461. .4546
FT                   /*tag=  d
FT                   /note= "indicated in the specification as an open reading
FT                   frame, encoding a protein with homologies to
FT                   pectinases"
FT   cds             5151. .6194
FT                   /*tag=  e
FT                   /note= "indicated in the specification as an open reading
FT                   frame, with homology to S. pombe thiamine
FT                   repressed gene"
FT   cds             5486. .5845
FT                   /*tag=  f
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             6386. .7414
FT                   /*tag=  g
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             6430. .7470
FT                   /*tag=  h
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             8051. .8917
FT                   /*tag=  i
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             9656. .10552
FT                   /*tag=  j
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             9996. .10364
FT                   /*tag=  k
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             10127. .10936
FT                   /*tag=  l
FT                   /note= "indicated in the specification as an open reading
FT                   frame, encoding a protein with homology to the
FT                   S. cerevisiae mitochondrial genome maintenance
FT                   protein (MGM101)"
FT   cds             11161. .11523
FT                   /*tag=  m
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             13441. .14517
FT                   /*tag=  n
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             14058. .14360
FT                   /*tag=  o
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             14989. .16119
FT                   /*tag=  p
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             16405. .17619
FT                   /*tag=  q
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             16491. .17336
FT                   /*tag=  r
FT                   /note= "indicated in the specification as an open reading
FT                   frame, encoding a protein with homology to the
FT                   S. cerevisiae ribosomal protein S7A (SCRPS7A)"
FT   cds             18429. .18797
FT                   /*tag=  s
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             18749. .21037
FT                   /*tag=  t
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             21763. .22488
FT                   /*tag=  u
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             22540. .23262
FT                   /*tag=  v
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
FT   cds             23667. .28607
FT                   /*tag=  w
FT                   /note= "indicated in the specification as an open reading
FT                   frame"
PN   WO9704106-A2.
PD   06-FEB-1997.
PF   15-JUL-1996; E03091.
PR   22-DEC-1995; EP-870137.
PR   14-JUL-1995; EP-870089.
PA   (INNO-) INNOGENETICS NV.
PI   Scarcez T,  Van Broekhoven A;
DR   WPI; 97-132645/12.
PT   New nucleic acid encoding Saccharomyces pectinase - useful in food
PT   processing, also its regulatory sequences are useful for expression
PT   of heterologous protein
PS   Example 1; Fig 2; 65pp; English.
CC   This sequence represents the insert from cosmid 109, containing
CC   Saccharomyces cerevisiae strain FY1679 DNA. This sequence contained the
CC   pectinase gene of the invention (see T80042). Pectinases are synthesised
CC   by plants and a variety of microorganisms, the best characterised
CC   pectinase being polygalacturonase. Pectinases play roles in the invasion
CC   of plant tissues by phytopathogens, the spoilage of fruits and
CC   vegetables, and also in their food processing and plant biological
CC   applications. Vectors containing the pectinase coding sequence are used
CC   for production of homologous or heterologous proteins in transformed host
CC   cells. The pectinase sequence may include regulatory and/or coding
CC   sequences, including the sequence encoding the pectinase signal peptide
CC   (which can be used in any recombinant system). The pectinase is useful in
CC   food processing (e.g. preparation of fruit juices), typical heterologous
CC   proteins that can be made are (poly)peptides of bacterial, animal or
CC   viral origin. Also, probes derived from the pectinase coding sequence can
CC   be used to detect the coding sequence and primers for amplification are
CC   used to isolate related genes.
SQ   Sequence   40875 BP;    12953 A;    7579 C;    8292 G;    12051 T;

  Query Match             4.4%;  Score 107.6;  DB 1;  Length 40875;
```

```
  Best Local Similarity   48.7%;  Pred. No. 2.7e-12;
  Matches  293;  Conservative    0;  Mismatches  309;  Indels    0;  Gaps    0;

Qy   1497 CGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCC 1556
          | ||| || || | ||     || | ||  || |  | ||  | . |  |  |  |  ||
Db  24579 CTACTCCTACCACCACTATCACTTCTACTACTTCTACCACATCGACTACTCCTACAACCT 24520

Qy   1557 CAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAG 1616
          | ||| |  |  || |  |         |  |  || || |  |  | ||  |  |  |
Db  24519 CTACTACTTCTACCACTCCTACAACTTCTACCACATCGACTACTCCTACAACCTCTACTA 24460

Qy   1617 TCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTA 1676
             |  | ||  |  ||| | ||||  |  |  | ||  || |  || |  |  || |||
Db  24459 CTTCTACCACTCCTACAACTTCTACCACATCGACTACTCCTACAACCTCTACTACTTCTA 24400

Qy   1677 CTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCG 1736
          | ||| || |  ||     || || | |||    |   | |  ||||| || || ||||
Db  24399 CCACTCCTACAACTTCTACCACATCGACTACTCCTACAACCTCTACTACTTCTACCACTC 24340

Qy   1737 CAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTG 1796
          |  |  || ||||  ||    |         |||   ||  |||  |  |  || |  |
Db  24339 CTACAACTTCTACAACTTCTACCACTCCTACAACTTCTACCACTCCTACAACTTCTACAA 24280

Qy   1797 CCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAA 1856
          || ||||| || | |  ||  |  ||    |  |  | |||||||| |   | ||   |
Db  24279 CCTCCACCACTTCTCAAACTTCCACCAAATCGACTACTCCTACTACTTCAAGCACATCGA 24220

Qy   1857 TCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCA 1916
             ||||     |  |  | |||||  || |  | ||  |  | || ||| |  |
Db  24219 CTACTCCTACAACTTCTACCACTCCTACAACTTCTACAACCTCTACCGCTCCTACAACTT 24160

Qy   1917 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCCCCCAACA 1976
            || ||    ||      |  ||||| || ||  | |||||||  |  |  || | |  |
Db  24159 CTACAACTTCTACAACCTCTACCACTTCTACCATTTCTACCGCTCCCACGACCTCTACAA 24100

Qy   1977 CCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATG 2036
          |  ||  |||| |     |  ||  |||| |  |  |  ||  |  | |   | |   |
Db  24099 CTTCAAGCACTTTCAGTACTTCTAGTGCCTCTGCATCCTCTGTCATTTCTACCACTGCTA 24040

Qy   2037 CACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAG 2096
          || ||   |||||   ||||    | ||||||  |   | || |  |  |  |  ||||
Db  24039 CAACCTCTACCACTTTTGCCAGTCTTACTACACCTGCCACATCTACGGCTTCTACTGACC 23980

Qy   2097 AT 2098
          ||
Db  23979 AT 23978
```

```
RESULT  10
Q51557/c
ID   Q51557 standard; DNA; 6530 BP.
AC   Q51557;
DT   17-MAY-1994  (first entry)
DE   Loricrin gene.
KW   Loricrin; gene; constitutive; inducible; vector; 5' flanking region;
KW   intron; 3' flanking region; TATA box; cap site; intron/exon boundary;
KW   DNA cassette; linker; restriction site; induction; human; psoriasis;
KW   epidermal cell; wound healing; ulcers; growth factor; antisense RNA;
KW   transforming growth factor beta; cytokine; skin; tracheal epithelia;
KW   cancer; E6; E7; papilloma virus; p53 protein; vaccine; oral epithelia;
KW   capsid protein; squamous epithelium; viral; oesophageal epithelia;
KW   vaginal epithelia; cornal epithelia; transduction; transplant; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   misc_rna        1. .1540
FT                   /*tag=  a
FT                   /note= "5' flanking region"
FT   intron          1587. .2677
FT                   /*tag=  b
FT                   /note= "Intron and intron/exon boundary"
FT   misc_rna        4384. .6530
FT                   /*tag=  c
FT                   /note= "3' flanking region"
```

```
PN   WO9322431-A.
PD   11-NOV-1993.
PF   28-APR-1993; U03993.
PR   30-APR-1992; US-876286.
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
PI   Greenhalgh DA,  Roop DR,  Rothnagel JA;
DR   WPI; 93-368790/46.
PT   Constitutive and inducible vectors based on loricrin or keratin
PT   K6 genes - for selective expression in the epidermis, e.g. for
PT   wound healing, treating psoriasis and skin cancer, etc.
PS   Claim 23; Page 26-32; 68pp; English.
CC   This sequence represents the loricrin gene and was used in the
CC   construction of the loricrin constitutive vector of the invention.
CC   The loricrin vector has a 5' flanking region of approx. 1.5 kb, an
CC   intron of approx. 1.1 kb and a 3' flanking region of approx. 1.5 kb.
CC   The 5' flanking region of the loricrin gene comprises a TATA box,
CC   cap site, first intron and intron/exon boundary, all in the correct
CC   order and position for expression of a nucleic acid cassette.  The
CC   loricrin vector also comprises a 3' flanking region of the loricrin
CC   gene and a linker with a unique restriction site at the position of
CC   the start and stop codons, joining the two flanking sequences and
CC   also having a position for insertion of the DNA cassette (see also
CC   Q51558).  Vectors such as this can be used for; (a) in vitro induction
CC   of human epidermal cells, especially for stimulating healing of
CC   wounds, surgical incisions and ulcers, (where the vector expresses a
CC   growth factor), (b) treating psoriasis, (the vector expressing
CC   transforming growth factor beta, a soluble cytokine or antisense
CC   RNA), (c) treating skin cancer, (the vector expressing antisense RNA
CC   of the E6 or E7 gene of human papilloma virus or normal p53 protein),
CC   (d) for vaccination (the vector expresses a viral capsid protein,
CC   especially of human papilloma virus), or (e) treating cancer of the
CC   squamous epithelium (of epidermal, oral, oesophageal, vaginal,
CC   tracheal or cornal epithelia, the vector expresses antisense RNA).
CC   Alternatively human epithelial cells are transduced ex vivo, then
CC   transplanted.
SQ   Sequence   6530 BP;    1635 A;    1507 C;    1696 G;    1692 T;
```

```
  Query Match             4.0%;  Score 99.4;  DB 1;  Length 6530;
  Best Local Similarity   48.5%;  Pred. No. 7.1e-11;
  Matches  308;  Conservative    0;  Mismatches  321;  Indels    6;  Gaps    1;

Qy   1497 CGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCC 1556
          | || || ||     |  | || ||    |  ||| |  | || ||   | |  | ||
Db   3595 CCACCGCAGCTGGAGCCTCCGCCCCCGGAGTAGCCGCCGCCGCAGCTGGAGCCACCGCTA 3536

Qy   1557 CAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAG 1616
           ||| ||| |     |    |   |   ||         | | ||||  | || ||| | |
Db   3535 GAACCGCCTCCGTAGCTCTGGCACTGATACTGTTGGGACGATCCGCCGCCGCCGCCACCG 3476

Qy   1617 TCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTA 1676
             | | ||  |||||  || | |  | ||| ||      ||| || |||| ||  |
Db   3475 GAGGAACCACCTCCGCAGCTCGTGCCGCCGCCACCGGAATAGCCGCCGCCGCCGCCGCAG 3416

Qy   1677 CTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCG 1736
          ||    |  || |    |    | || | | |||||    | |  | ||||| ||||||
Db   3415 CTGGATCCACCGCCACCTGAGTAGCCGCCGCCACAGCTAGAGCCGCCGCTGCTTCCACCG 3356

Qy   1737 CAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTG 1796
          | |||   |   | ||   | | |      | | | | | |||   ||  |       ||
Db   3355 CCGCCGGAATAGCCGCCTCCGCAGCTGGAGCCACCGCCGCCACCAGAGTACT------TG 3302

Qy   1797 CCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAA 1856
          || || |||   || || |||| ||| ||   || ||| |  | |  |  | | | | |
Db   3301 CCGCCTCCGGAACCGCCGCCGCAGCCACCGGAGCTGCCGCCTCCAGAGGAGCCTCCGCAG 3242

Qy   1857 TCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCA 1916
          |   | |  | | | | | | |  |  | || ||     |  |||| ||  |  ||
Db   3241 TAGGAGCCGCCGCCTCCTGATCCGCCTCCGCAGCTGGAGCCGCCGCCGCCGCCCCCGGAG 3182

Qy   1917 ACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACA 1976
            || | | | | |    | || | || ||| | ||  ||| || || || | ||  ||
Db   3181 GAGCCGCCGCCGCAGCTAGAGCCGCCGCCGCCACCACCTCCGGAGTACTTGACGCCCCCA 3122
```

```
Qy   1977 CCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATG 2036
          ||| || | |     |  |||| ||||   ||||  | |      || ||| |||    |
Db   3121 CCAGAGCCGCCACCGCCTCCACAGCTGGAGCCACCCCCGGAGTAGCTGCCGCCGCCGCAG 3062

Qy   2037 CACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAG 2096
          |      | || | ||  ||   ||| | ||  ||   ||   ||| || | |
Db   3061 CTGGAGCCGCCGCCGCCTCCGGAGTAGCCACCTCCGCAGCTGGAGCCGCCGCCGCCCCCA 3002

Qy   2097 ATGATGACATCGACCCCCCTCTCCCCAACGACCCC 2131
            |  | || || | |  ||    ||| |  | ||
Db   3001 GAGTAGCCACCGCCGCAGCTAGAGCCACCACCGCC 2967
```

RESULT 11
Q06483
ID Q06483 standard; cDNA; 502 BP.
AC Q06483;
DT 22-FEB-1991 (first entry)
DE Intestinal mucin clone SMUC 40.
KW Epithelial cancer; cystic fibrosis; ulcerative colitis;
KW glycosyl transferase; familial polyposis coli;
KW glycosylation patterns; ss.
OS Homo sapiens.

FH	Key	Location/Qualifiers
FT	repeat_unit	9. .77
FT		/*tag= a
FT		/number=1
FT	repeat_unit	78. .146
FT		/*tag= b
FT		/number=2
FT	repeat_unit	147. .215
FT		/*tag= c
FT		/number=3
FT	repeat_unit	216. .284
FT		/*tag= d
FT		/number=4
FT	repeat_unit	285. .353
FT		/*tag= e
FT		/number=5
FT	repeat_unit	354. .422
FT		/*tag= f
FT		/number=6
FT	repeat_unit	423. .490
FT		/*tag= g
FT		/number=7
FT	cds	3. .500
FT		/*tag= h

PN WO9012892-A.
PD 01-NOV-1990.
PF 25-JUL-1989; U03206.
PR 14-APR-1989; US-338710.
PA (REGC) University of California.
PI Kim YS; Gum JR
DR WPI; 90-348495/46.
DR P-PSDB; R07670.
PT Human intestinal mucin DNA, polypeptides and antibodies - used in
PT diagnosis and treatment of eg. cancers, cystic fibrosis and
PT colitis.
PS Disclosure; Fig 4A; 45pp; English.
CC The clone encodes intestinal mucin. It was isolated from a cDNA
CC library using antibody probes. The DNA sequence allows
CC determination of the primary amino acid structure, difficult to
CC characterise biochemically because the protein backbone is so
CC heavily glycosylated and conditions required to remove the
CC carbohydrate tend to result in breakage of the protein. The DNA
CC has been used to characterise expression and genomic organisation
CC of the hIM family. Prods. of the DNA can be used to produce Abs
CC useful for detecting hIM and determining glycosylation patterns
CC or glycosyltransferase activity. They can be used for the early
CC detection and differential diagnosis of cancers esp. epithelial
CC cancer, and also in diseases with altered intestinal mucin prodn.
CC such as cystic fibrosis and colitis.
CC See also Q06484-89.
SQ Sequence 502 BP; 162 A; 262 C; 52 G; 26 T;

Query Match 4.0%; Score 98.8; DB 1; Length 502;
Best Local Similarity 53.2%; Pred. No. 5.7e-11;
Matches 233; Conservative 0; Mismatches 202; Indels 3; Gaps 1;

```
Qy    308 ACCCTGCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACA 367
          ||||   |   | | ||  ||| |||           |  ||||| |     || | ||
Db     45 ACCCCAACCCCAACACCCACTGGAACACAGACCCTAACCCCAACACCCATCACCACCACC 104

Qy    368 ACCACCCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACC 427
          || ||   | |||||||| | ||  |    | | |    ||  | ||   | | || | |
Db    105 ACTACGGTGACCCCAACCCCTACACCCACCGGCACACAGACCCCAACATCGACACCCATC 164

Qy    428 ACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACC 487
          |  ||||||    |    ||| | || |   | || |||  ||| ||| |||||||| ||
Db    165 AGCACCACCACTACGGTGACCCCAACACCAACACCCACCGGCACACAGACCCCAACCACG 224

Qy    488 ACTCAGGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACA 547
          || |    | | || || || ||   | |||||||| ||||  |    | | |    ||
Db    225 ACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACC 284

Qy    548 ACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACT 607
          | |||| | | || | |||||||    || | ||    ||||   |||||||  ||||
Db    285 CCAACCCGGACACCCATCACCACC---ACCACCACGGTGACCCCAACCCCAACACCCACC 341

Qy    608 CAGGCCCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCA 667
            | |  ||  ||| | ||    |    || ||| |  || |    |||||  || ||
Db    342 GGCACACAGACCCCAACCCCCACACCCATCACCACCACCACCACCGTGACCCCAACCCCA 401

Qy    668 ACTACCCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTT 727
          ||  || | | | ||   |||||  |  ||||| ||  | |  | |  ||  | |
Db    402 ACACCCACCGGCACACAGACCCCAACATCGACACCCATCACCACCAACACTACGGTGACC 461

Qy    728 CCTGCACCCACTTCAGCC 745
          ||  | || ||  || ||
Db    462 CCAACCCCAACACCAACC 479
```

RESULT 12
N91235
ID N91235 standard; DNA; 1368 BP.
AC N91235;
DT 07-FEB-1990 (first entry)
DE DNA sequence of preprospasmolysin.
KW Preprospasmolysin; Xenopus laevis.
OS Xenopus laevis

FH	Key	Location/Qualifiers
FT	cds	28. .1230
FT		/*tag= a

PN DE3808456-A.
PD 28-SEP-1989.
PF 14-MAR-1988; 808456.
PR 14-MAR-1988; DE-808456.
PA (PLAC) Max-Plank Ges. Wissensch.
PI Weickmann H, Fincke K, Weickmann F, Huber B, Liska H, Prechtel J;
DR WPI; 89-286097/40. P-PSDB; P91941.
PT New polypeptide preprospasmolysin and its fragments- derived from Xenopus
PT laevis, with spasmolytic and neurotransmitter activities.
PS Disclosure; Fig. 2; 8pp; german.
CC The DNA or mRNA corresp. to the peptides are derived from the skin of
CC Xenopus laevis. The DNA sequence is inserted into a vector and used to
CC transform eukaryotic cells, esp. COS cells, or the DNA is converted into
CC mRNA and microinjected into eukaryotic cells, esp. X. laevis oocytes.
SQ Sequence 1368 BP; 511 A; 332 C; 234 G; 291 T;

Query Match 3.9%; Score 95.6; DB 1; Length 1368;
Best Local Similarity 48.8%; Pred. No. 2.8e-10;
Matches 291; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

```
Qy   1515 CTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTCCA 1574
Db    371 CAGTTATAACAACTCCAGCACCAGATACAACTACAGCTTCAACAACTGCAGAAACAACTA 430

Qy   1575 CCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAA 1634
Db    431 CAGTTCCAACAACTCCAGAAACAACTACAGTTCCAACAACTCCAGAAACAACTACAGTTC 490

Qy   1635 TTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAAT 1694
Db    491 CAACAACTCCAGAAACAACTACAGTTCCAACAACTCCAGAAACAACTACAGTTCCAACAA 550

Qy   1695 CCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTG 1754
Db    551 CTCCAGAAACAACTACAGTTCCAACAACTCCAGAAACAACTACAGTTCCAACAACTCCAG 610

Qy   1755 TCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCA 1814
Db    611 AAACAACTACAGTTCCAACAACTCCAGAAACAACTACAGTTCCAACAACTCCAGAAACAA 670

Qy   1815 CCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACAT 1874
Db    671 CTACAGTTCCAACAACTCCAGAAACAACTACAGTTCCAACAACTCCAGAAACAACTACAG 730

Qy   1875 CACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTAC 1934
Db    731 CTTCAACAACTGCAGAAACAACTACAGTTCCAACAACTCCAGAAAC------AACTACAG 784

Qy   1935 CACCCACTGCTGCCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCAC 1994
Db    785 AACCGACTACAACACCAACAACAGATACTACACCACCAACACTACCACCAACACCAGAAA 844

Qy   1995 CCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCAACACCACAGCTG 2054
Db    845 CAACTACAGAAACAACTACAGAAACAACTACAGAAACAACTACAGAAACAACTACAGAAA 904

Qy   2055 CCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGAC 2110
Db    905 CAACTACAGAAACAACTACAGCTCCACCACCAGAATGTGCAGCTGACAGAGTGGAC 960
```

RESULT 13
Q06486
ID Q06486 standard; cDNA; 860 BP.
AC Q06486;
DT 22-FEB-1991 (first entry)
DE Intestinal mucin clone SMUC 53.
KW Epithelial cancer; cystic fibrosis; ulcerative colitis;
KW glycosyl transferase; familial polyposis coli;
KW glycosylation patterns; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT repeat_unit 62. .130
FT /*tag= a
FT /number=2
FT repeat_unit 131. .199
FT /*tag= b
FT /number=3
FT repeat_unit 200. .258
FT /*tag= c
FT /number=4
FT repeat_unit 259. .327
FT /*tag= d
FT /number=5
FT cds 2. .860
FT /*tag= e
PN WO9012892-A.
PD 01-NOV-1990.
PF 25-JUL-1989; U03206.
PR 14-APR-1989; US-338710.
PA (REGC) University of California.
PI Kim YS; Gum JR
DR WPI; 90-348495/46.
DR P-PSDB; R07673.
PT Human intestinal mucin DNA, polypeptides and antibodies - used in
PT diagnosis and treatment of eg. cancers, cystic fibrosis and
PT colitis.
PS Disclosure; Fig 4D; 45pp; English.
CC The clone encodes intestinal mucin. It was isolated from a cDNA
CC library using antibody probes. The DNA sequence allows
CC determination of the primary amino acid structure, difficult to
CC characterise biochemically because the protein backbone is so
CC heavily glycosylated and conditions required to remove the
CC carbohydrate tend to result in breakage of the protein. The DNA
CC has been used to characterise expression and genomic organisation
CC of the hIM family. Prods. of the DNA can be used to produce Abs
CC useful for detecting hIM and determining glycosylation patterns
CC or glycosyltransferase activity. They can be used for the early
CC detection and differential diagnosis of cancers esp. epithelial
CC cancer, and also in diseases with altered intestinal mucin prodn.
CC such as cystic fibrosis and colitis.
CC See also Q06483-89.
SQ Sequence 860 BP; 255 A; 364 C; 94 G; 147 T;

Query Match 3.8%; Score 93; DB 1; Length 860;
Best Local Similarity 57.1%; Pred. No. 8.2e-10;
Matches 190; Conservative 0; Mismatches 140; Indels 3; Gaps 1;

```
Qy    350 ACAACTCAGGCCCCAACAACCACCCAGGCCCCAACCACAACTACTCAGGCCCCTACTACA 409
Db      2 ACACCCATCACCACCACCACCACCGTGACACCAACCCCAACACCCACCGGCACACAGACC 61

Qy    410 ACCACCCAGGCCCCAACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACCACC 469
Db     62 CCAACCATGACACCCATCACCACCACCACCACGGTGACCCCAACCCCAACACCCACCGGC 121

Qy    470 ACCCAGGCCCCAACTACCACTCAGGCCCCTACTACTACCACTCAGGCCCCAACC---ACA 526
Db    122 ACACAGACCCCAACCATGACACCCATCACCACCACCACCACCGTGACCCCAACCCCAACA 181

Qy    527 ACCACTCAGGCCCCTACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACC 586
Db    182 CCCACTGGCACACAGGCCCCAACCCCAACAGCCATCACCACCACCACTACGGTGACCCCA 241

Qy    587 ACCCAGGCCCCAACTACCACTCAGGCCCCAACTACAATCACCCAGGCTGCAACTACCCCG 646
Db    242 ACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACCACG 301

Qy    647 GCCGCAACTACCCCGGCCGCAACTACCCCGGCC 679
Db    302 GTGACCCCAACCCCAACACCCAACCCCCTGGTC 334
```

RESULT 14
T13030
ID T13030 standard; DNA; 1984 BP.
AC T13030;
DT 27-MAY-1996 (first entry)
DE Cotton fibre-specific H6 gene and promoter.
KW Cotton; fibre; promoter; transgenic plant; crop improvement; ds.
OS Gossypium barbadense strain Sea Island.
FH Key Location/Qualifiers
FT promoter 1. .250
FT /*tag= a
FT 5'utr 251. .320
FT /*tag= b
FT cds 321. .1547
FT /*tag= c
FT /*note= "the CDS includes a 583 bp intron
FT (location unspecified)"
FT 3'utr 1548. .1984
FT /*tag= d
PN US5495070-A.
PD 27-FEB-1996.
PF 04-OCT-1988; 253243.

```
PR   04-OCT-1988; US-253243.
PR   21-NOV-1990; US-617239.
PR   18-MAY-1992; US-885970.
PA   (CETU ) AGRACETUS INC.
PI   John M;
DR   WPI; 96-139095/14.
PT   New isolated fibre-specific promoters - used for introducing
PT   altered fibre-specific characteristics into plants, partic. cotton.
PS   Claim 1; Column 59-62; 48pp; English.
CC   The cotton var. Sea Island H6 gene and promoter (T13030) were
CC   identified in genomic clone EMBL-S1-H6-4 isolated by screening a
CC   genomic library prepd. in lambda EMBL with cotton fibre-specific
CC   cDNA clone H6 (see T13034).  H6 and other (see T13025-29, T13031-32
CC   and T13052-53) fibre-specific promoters can be utilised in the
CC   construction of vectors used for the introduction of altered fibre-
CC   specific characteristics into plants, partic. cotton.  They can be
CC   used to modulate the synthesis of fibre proteins or to introduce
CC   non-fibre proteins into fibre in a tissue-specific manner.
SQ   Sequence   1984 BP;    565 A;    470 C;    281 G;    668 T;


  Query Match             3.8%;  Score 93;  DB 1;  Length 1984;
  Best Local Similarity   53.1%;  Pred. No. 9.6e-10;
  Matches  198;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy   1593 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1652
           |||||||| || |  | ||  |    || ||| | ||  |     |  | ||  | || |
Db    424 CGCCACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 483

Qy   1653 CCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1712
           ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    484 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 543

Qy   1713 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1772
             | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||
Db    544 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 603

Qy   1773 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1832
           |||   |  |  | ||| |  || | ||  | ||| | ||  |  |||||  |  |  ||
Db    604 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 663

Qy   1833 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1892
           |  || || |  ||  | |       ||| || |   | || | ||| | ||| || | |
Db    664 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 723

Qy   1893 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1952
           |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    724 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 783

Qy   1953 CTACCGCAGCACC 1965
           |      | ||||
Db    784 CAGTACAAACACC 796


RESULT  15
T30250
ID   T30250 standard; DNA; 1985 BP.
AC   T30250;
DT   16-DEC-1996  (first entry)
DE   Cotton fibre clone SKSIE6-H6-RI-derived promoter contg. sequence.
KW   Cotton fibre; promoter; differential screening; leaf; ovule; root;
KW   flower; PCR; polymerase chain reaction; homology; transgenic plant; ds.
OS   Gossypium barbadense.
FH   Key             Location/Qualifiers
FT   promoter        1. .250
FT                   /*tag=  a
FT                   /note= "contains promoter sequence (claimed)"
FT   misc_feature    251. .321
FT                   /*tag=  b
FT                   /note= "5' non-coding sequence"
FT   cds             322. .1547
FT                   /*tag=  c
FT                   /note= "contains a 583 bp intron"
FT   misc_feature    1548. .1984
FT                   /*tag=  d
FT                   /note= "3' non-coding sequence"
PN   US5521078-A.
PD   28-MAY-1996.
PF   04-OCT-1988; 253243.
PR   04-OCT-1988; US-253243.
PR   21-NOV-1990; US-617239.
PR   18-MAY-1992; US-885970.
PR   19-OCT-1994; US-298687.
PA   (CETU ) AGRACETUS INC.
PI   John M;
DR   WPI; 96-267794/27.
PT   Isolation of fibre-specific cotton promoter sequences - using
PT   selected DNA probes to screen genomic DNA fragments, for production
PT   of cotton fibres with improved characteristics
PS   Claim 1; Column 59-62; 46pp; English.
CC   Cotton fibre cell-specific promoter sequences were isolated by
CC   differential screening of a cotton plant cDNA library.  Of 4788 clones
CC   from a 10 day cell library screened with leaf cDNAs, 800 clones not
CC   present in the leaf were screened with cDNAs from
CC   ovule, root and flower mRNAs and resulted in 79 clones isolated.  PCR
CC   analysis was then used to remove cross-hybridising clones.  This resulted
CC   in the isolation of 18 cDNA clones specifically expressed in cotton fibre
CC   cells (T30242-4 and T30253-67). These cDNAs were then used to screen for
CC   homologous genomic sequences (T30245-53 and T30268) in order to obtain
CC   the corresp. promoter sequences.
CC   This genomic clone is derived from the 1.9 kb SalI-EcoRI fragment of the
CC   sub-clone designated SKSIE6-H6-RI.  This forms part of the genomic clone
CC   designated as EMBL-SI-H6-4 which contains a 13 kb MboI insert.
CC   The promoters isolated from the fibre cell-specific clones can be used to
CC   generate transgenic cotton plants and lines producing fibres having
CC   altered quantity and quality.
SQ   Sequence   1985 BP;    565 A;    470 C;    282 G;    668 T;


  Query Match             3.8%;  Score 93;  DB 1;  Length 1985;
  Best Local Similarity   53.1%;  Pred. No. 9.6e-10;
  Matches  198;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy   1593 CGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCA 1652
           |||||||| || |  | ||  |    || ||| | ||  |     |  | ||  | || |
Db    424 CGCCACCCACACCACCGGCTTCTACTCCTCCTCCGACGACTCAAGCACCGCCTACACCAA 483

Qy   1653 CCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAG 1712
           ||||  | ||  |  | | | ||||| || || || ||  ||| ||  || | || ||||
Db    484 CCGCCACTCCGCCACCGGTTTCTACTCCTCCTCCCACTTCATCACCGCCCCCAGTGACAG 543

Qy   1713 TACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCA 1772
             | ||  |  | | | | ||  |  | || | | || ||   |||| ||  |||
Db    544 CTTCTCCACCCCCAGTTTCAACTCCTCCACCCAGTTCTCCTCCTCCTGCAACTCCACCAC 603

Qy   1773 CTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAG 1832
           |||   |  |  | ||| |  || | ||  | ||| | ||  |  |||||  |  |  ||
Db    604 CTGCTTCTCCTCCTCCTGCAACTCCACCTCCAGCTTCTCCACCTCCTGCCACTCCTCCAC 663

Qy   1833 CCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCC 1892
           |  || || |  ||  | |       ||| || |   | || | ||| | ||| || | |
Db    664 CAGCTTCTCCACCTCCCGCCACTCCACCACCTGCAACCCCACCGCCAGCAACTCCTCCTC 723

Qy   1893 CCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTA 1952
           |  ||||| || |||| || ||     |  |     | |   ||  | | |||  | ||
Db    724 CTGCTACCCCACCACCAGCTCCATTGGCTTCTCCTCCAGCCACAGTCCCAGCTATCTCTC 783

Qy   1953 CTACCGCAGCACC 1965
           |      | ||||
Db    784 CAGTACAAACACC 796


Search completed: July 18, 1999, 11:07:49
Job time: 2139 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 05:08:34 ; Search time 2471.47 Seconds
(without alignments)
3553.707 Million cell updates/sec

Title: US-09-294-663-1
Perfect score: 2455
Sequence: 1 GTAACGTTAAGTGAAAAGAA.....AAAAAAAAAAAAAAAAAAAA 2455

Scoring table: IDENTITY_NUC

Searched: 808301 seqs, 1788773984 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl1:*
9: gb_pl2:*
10: gb_pr1:*
11: gb_pr2:*
12: gb_pr3:*
13: gb_ro:*
14: gb_st:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: gb_htg:*
19: em_ba:*
20: em_fun:*
21: em_hum1:*
22: em_hum2:*
23: em_in:*
24: em_om:*
25: em_or:*
26: em_ov:*
27: em_pat:*
28: em_ph:*
29: em_pl:*
30: em_ro:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: em_htg:*
35: em_sts:*
36: gb_ba1:*
37: gb_ba2:*
38: gb_pl1:*
39: gb_pl2:*
40: gb_pr1:*
41: gb_pr2:*
42: gb_pr3:*
43: gb_sts:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2455	100.0	2455	3	TNAF000605	AF000605 Trichoplu
	2	2354.4	95.9	2821	3	TNAF000606	AF000606 Trichoplu
	3	193	7.9	15720	11	HUMMUC2X	L21998 Homo sapien
	4	193	7.9	15720	41	HUMMUC2X	L21998 Homo sapien
c	5	151.8	6.2	185994	18	AC002042	AC002042 *** SEQUE
c	6	151.6	6.2	10596	6	I25041	I25041 Sequence 15
c	7	151.6	6.2	10596	6	I30503	I30503 Sequence 15
c	8	151.6	6.2	5452	15	U02454	U02454 Cloning vec
c	9	151.6	6.2	10850	15	U02455	U02455 Cloning vec
c	10	151.6	6.2	10737	15	XXU02428	U02428 Cloning vec
c	11	151.6	6.2	172281	17	EBV	V01555 Epstein-Bar
c	12	151.6	6.2	184113	17	HS4B958RAJ	M80517 Epstein-Bar
c	13	151.6	6.2	1150	17	HS4ULIR3	J02079 epstein-bar
	14	145.6	5.9	3170	17	D88733	D88733 Equine herp
	15	140	5.7	3173	17	D88734	D88734 Equine herp
	16	132.8	5.4	2070	5	BSU43200	U43200 Boreogadus
c	17	130	5.3	2479	13	RNU87960	U87960 Rattus norv
c	18	127	5.2	3941	13	RNLAR2	X83546 R.norvegicu
	19	125.8	5.1	33359	3	CEF55B11	Z83318 Caenorhabdi
c	20	124.4	5.1	28291	3	CEF56H9	Z74473 Caenorhabdi
	21	121.6	5.0	5719	13	MMCATS1	AF051726 Mus muscu
	22	120.8	4.9	1046	5	PSUAFA02	M63477 P.americanu
c	23	118.8	4.8	4635	12	HSA6805	AJ006805 Homo sapi
c	24	118.8	4.8	4635	42	HSA6805	AJ006805 Homo sapi
	25	118.4	4.8	1720	10	HUMMUC2A	M74027 Human mucin
	26	118.4	4.8	1720	40	HUMMUC2A	M74027 Human mucin
c	27	117.6	4.8	174428	12	AC005821	AC005821 Homo sapi
c	28	117.6	4.8	174428	42	AC005821	AC005821 Homo sapi
	29	116.4	4.7	176966	18	AC005139	AC005139 *** SEQUE
	30	116	4.7	1988	11	AC002494	AC002494 Human PCR
	31	116	4.7	1988	41	AC002494	AC002494 Human PCR
	32	115.2	4.7	3811	10	HUMIMUCA	M94131 Human mucin
	33	115.2	4.7	3811	40	HUMIMUCA	M94131 Human mucin
c	34	114.8	4.7	186324	18	AC005505	AC005505 *** SEQUE
c	35	114	4.6	123110	12	AC005383	AC005383 Homo sapi
c	36	114	4.6	123110	42	AC005383	AC005383 Homo sapi
c	37	113.8	4.6	11343	13	MMDSPPG	AJ002141 Mus muscu
c	38	113.8	4.6	4420	13	U67916	U67916 Mus musculu
c	39	113	4.6	2793	6	AR009990	AR009990 Sequence
c	40	113	4.6	2793	6	I95876	I95876 Sequence 1
c	41	111.4	4.5	3060	3	AF027735	AF027735 Nephila c
c	42	111	4.5	48000	18	AC005930	AC005930 *** SEQUE
c	43	108.6	4.4	3489	17	KSU52064	U52064 Kaposi's sa
c	44	108.4	4.4	1569	10	GIBINVOL	M35447 Hylobates l
c	45	108.4	4.4	1569	40	GIBINVOL	M35447 Hylobates l

ALIGNMENTS

RESULT 1
TNAF000605
LOCUS TNAF000605 2455 bp mRNA INV 28-JUN-1997
DEFINITION Trichoplusia ni insect intestinal mucin IIM14 mRNA, complete cds.
ACCESSION AF000605
NID g2224918
KEYWORDS .
SOURCE cabbage looper.
ORGANISM Trichoplusia ni
Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Noctuoidea;
Noctuidae; Trichoplusia.
REFERENCE 1 (bases 1 to 2455)
AUTHORS Wang,P. and Granados,R.R.
TITLE Molecular cloning and sequencing of a novel invertebrate intestinal mucin cDNA
JOURNAL J. Biol. Chem. 272 (26), 16663-16669 (1997)
MEDLINE 97341213
REFERENCE 2 (bases 1 to 2455)
AUTHORS Wang,P. and Granados,R.R.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1997) Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA

```
FEATURES             Location/Qualifiers
     source          1. .2455
                     /organism="Trichoplusia ni"
                     /db_xref="taxon:7111"
     CDS             38. .2404
                     /codon_start=1
                     /product="insect intestinal mucin IIM14"
                     /db_xref="PID:g2224919"
                     /translation="MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWL
                     LPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQA
                     PATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQA
                     PTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTTP
                     AATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGY
                     TFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPA
                     NFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGES
                     EEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA
                     RPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPT
                     AAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPT
                     AAPTTAVPEIPITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP
                     APNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPA
                     TTPEDDDIDPPLPNDPINPCVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNP
                     TTKTCDFACNVGCVRSNIQMSESYEGVQVFIPWNKLDEDIRQALNFEL"
BASE COUNT       592 a    961 c    479 g    423 t
ORIGIN
```

Query Match 100.0%; Score 2455; DB 3; Length 2455;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GTAACGTTAAGTGAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTAACGTTAAGTGAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCT 60

Qy      61 GACGGCCCTCGGGCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GACGGCCCTCGGGCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGC 120

Qy     121 TCTCCACGAGCCGCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCTCCACGAGCCGCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGA 180

Qy     181 ATACGACTGCACCAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATACGACTGCACCAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGA 240

Qy     241 CTGTGCTCCTGGTACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTGTGCTCCTGGTACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGC 300

Qy     301 CGGATGCACCCTGCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CGGATGCACCCTGCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGC 360

Qy     361 CCCAACAACCACCCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCCAACAACCACCCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGC 420

Qy     421 CCCAACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CCCAACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCC 480

Qy     481 AACTACCACTCAGGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AACTACCACTCAGGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCC 540

Qy     541 TACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAAC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TACCACAACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAAC 600

Qy     601 TACCACTCAGGCCCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TACCACTCAGGCCCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCC 660

Qy     661 GGCCGCAACTACCCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GGCCGCAACTACCCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCC 720

Qy     721 AGGTGTTCCTGCACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AGGTGTTCCTGCACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAA 780

Qy     781 TGGTTGCCCAGCTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TGGTTGCCCAGCTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCT 840

Qy     841 CTTCTACCAGTGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTTCTACCAGTGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTT 900

Qy     901 CAACCCCTACGTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 CAACCCCTACGTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAG 960

Qy     961 CCCCGCACCCCCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CCCCGCACCCCCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCT 1020

Qy    1021 CGACAATGGATGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CGACAATGGATGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTG 1080

Qy    1081 CGACAAGTATTACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCAC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CGACAAGTATTACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCAC 1140

Qy    1141 CCACTTCAGTTTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CCACTTCAGTTTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCC 1200

Qy    1201 CGGCGGCGAGAGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CGGCGGCGAGAGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCC 1260

Qy    1261 CACGGAACCCATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGA 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CACGGAACCCATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGA 1320

Qy    1321 CCACCTCCTCCCCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CCACCTCCTCCCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGAC 1380

Qy    1381 TATCGCAAGACCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TATCGCAAGACCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTC 1440

Qy    1441 TCCTGTGACCGCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGAC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TCCTGTGACCGCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGAC 1500

Qy    1501 TGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAAC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 TGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAAC 1560

Qy    1561 TGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCC 1620

Qy    1621 ACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTAC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 ACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTAC 1680

Qy    1681 TGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 TGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGC 1740
```

```
Qy   1741 CCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCC 1800

Qy   1801 CACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCC 1860

Qy   1861 AACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACAC 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 AACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACAC 1920

Qy   1921 CACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACACCAC 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACACCAC 1980

Qy   1981 AGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACC 2040

Qy   2041 CAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGA 2100

Qy   2101 TGACATCGACCCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGT 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 TGACATCGACCCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGT 2160

Qy   2161 TTTGCCCTGGGCTCACGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TTTGCCCTGGGCTCACGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGT 2220

Qy   2221 CCTGGTGGTTTGCTCTGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGC 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CCTGGTGGTTTGCTCTGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGC 2280

Qy   2281 TTGCAACGTCGGTTGCGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCA 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TTGCAACGTCGGTTGCGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCA 2340

Qy   2341 GGTCTTCATCCCATGGAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTT 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GGTCTTCATCCCATGGAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTT 2400

Qy   2401 GTAAACCTACTTAAATTAATGAAGGTTTTGTTTTAAAAAAAAAAAAAAAAAAAAAA 2455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GTAAACCTACTTAAATTAATGAAGGTTTTGTTTTAAAAAAAAAAAAAAAAAAAAAA 2455
```

```
RESULT   2
TNAF000606
LOCUS       TNAF000606   2821 bp    mRNA            INV       28-JUN-1997
DEFINITION  Trichoplusia ni insect intestinal mucin IIM22 mRNA, complete cds.
ACCESSION   AF000606
NID         g2224920
KEYWORDS    .
SOURCE      cabbage looper.
  ORGANISM  Trichoplusia ni
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
            Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Noctuoidea;
            Noctuidae; Trichoplusia.
REFERENCE   1  (bases 1 to 2821)
  AUTHORS   Wang,P. and Granados,R.R.
  TITLE     Molecular cloning and sequencing of a novel invertebrate intestinal
            mucin cDNA
  JOURNAL   J. Biol. Chem. 272 (26), 16663-16669 (1997)
  MEDLINE   97341213
REFERENCE   2  (bases 1 to 2821)
  AUTHORS   Wang,P. and Granados,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1997) Boyce Thompson Institute, Cornell
            University, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .2821
                     /organism="Trichoplusia ni"
                     /db_xref="taxon:7111"
     CDS             26. .2449
                     /codon_start=1
                     /product="insect intestinal mucin IIM22"
                     /db_xref="PID:g2224921"
                     /translation="MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWL
                     LPHEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQA
                     PATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQA
                     PTTTTQAPTTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTP
                     AATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGY
                     TFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPA
                     NFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGES
                     EEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA
                     RPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPT
                     AAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPT
                     AAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP
                     APNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPT
                     VAPAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHADCDKYWV
                     CDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMSESYEGVQVFIPWNKLDED
                     IRQALNFEL"
BASE COUNT      710 a    1030 c     529 g     552 t
ORIGIN
```

Query Match 95.9%; Score 2354.4; DB 3; Length 2821;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 2422; Conservative 0; Mismatches 1; Indels 57; Gaps 1;

```
Qy     13 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 72
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAAAAGAATAACCAGCGAACAAGTTATGATAAAGACCCTCCTATTCCTGACGGCCCTCGG 60

Qy     73 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTCGTCGCCGCGCGTCCTGAAGTCAGCGACGCGGAGAAGAACCCCGCTCTCCACGAGCC 120

Qy    133 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 180

Qy    193 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTCATCGCACCGAGAGACTGTGCTCCTGG 240

Qy    253 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCT 300

Qy    313 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCAGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCAC 360

Qy    373 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCAGGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAAC 420

Qy    433 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CACCCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCA 480

Qy    493 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGCCCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCAC 540

Qy    553 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CCAGGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGC 600

Qy    613 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CCCAACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTAC 660
```

```
Qy    673 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CCCGGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGC 720

Qy    733 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ACCCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAGC 780

Qy    793 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAGTG 840

Qy    853 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900

Qy    913 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCCCC 960

Qy    973 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1032
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGATG 1020

Qy   1033 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTATTA 1080

Qy   1093 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGTTT 1140

Qy   1153 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAGAG 1200

Qy   1213 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260

Qy   1273 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTCCC 1320

Qy   1333 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380

Qy   1393 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACCGC 1440

Qy   1453 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCGAC 1500

Qy   1513 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCCTC 1560

Qy   1573 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCCCACCTACAACCGC 1620

Qy   1633 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCTGCCCCTGA 1680

Qy   1693 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 ATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCTACTACTGC 1740

Qy   1753 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 TGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACCGCTGCCCC 1800

Qy   1813 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACTACTGTCAC 1860

Qy   1873 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGT 1920

Qy   1933 ACCACCCACTGCT---------------------------------------------- 1946
          |||||||||||||
Db   1921 ACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCTCCCAACACCACAGTGACTGCACC 1980

Qy   1946 ----------GCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 1995
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 ACCCACCGCAGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTCACTGTACCACC 2040

Qy   1996 CACTGCTGCCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCAACACCACAGCTGC 2055
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   2041 CACTGCTGCCCCCCACTGCAGCTCCCCCTACCGTCGCACCTGCACCCAACACCACAGCTGC 2100

Qy   2056 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2115
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 CCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAGATGATGACATCGACCCCCC 2160

Qy   2116 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TCTCCCCAACGACCCCATCAACCCTTGCGTTGAAGAATGCAACGTTTTGCCCTGGGCTCA 2220

Qy   2176 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CGCTGACTGCGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCCTGGTGGTTTGCTC 2280

Qy   2236 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG 2340

Qy   2296 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CGTGAGGAGCAACATTCAGATGTCTGAAAGCTACGAAGGAGTCCAGGTCTTCATCCCATG 2400

Qy   2356 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GAACAAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTACTTAAA 2460

Qy   2416 TTAATGAAGGTTTTGTTTTA 2435
          ||||||||||||||||||||
Db   2461 TTAATGAAGGTTTTGTTTTA 2480
```

RESULT 3
HUMMUC2X
LOCUS HUMMUC2X 15720 bp mRNA PRI 09-JAN-1998
DEFINITION Homo sapiens intestinal mucin (MUC2) mRNA, complete cds.
ACCESSION L21998
NID g454153
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 15720)
AUTHORS Gum,J.R.Jr. ., Hicks,J.W., Toribara,N.W., Kim,Y.S. and Siddiki,B.
TITLE Molecular cloning of human intestinal mucin (MUC2) cDNA: Identification of the amino terminal and overall sequence similarity to pre-pro-von Willebrand factor
JOURNAL J. Biol. Chem. 269, 2440-2446 (1994)
MEDLINE 94132002
REFERENCE 2 (bases 1 to 15720)
AUTHORS Eckhardt,A.E., Timpte,C.S., DeLuca,A.W. and Hill,R.L.

```
TITLE       The complete cDNA sequence and structural polymorphism of the
            polypeptide chain of porcine submaxillary mucin
  JOURNAL   J. Biol. Chem. 272 (52), 33204-33210 (1997)
  MEDLINE   98070526
FEATURES             Location/Qualifiers
     source          1. .15720
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="intestine"
     CDS             28. .15567
                     /codon_start=1
                     /product="mucin"
                     /db_xref="PID:g454154"
                     /translation="MGLPLARLAAVCLALSLAGGSELQTEGRTRYHGRNVCSTWGNFH
                     YKTFDGDVFRFPGLCDYNFASDCRGSYKEFAVHLKRGPGQAEAPAGVESILLTIKDDT
                     IYLTRHLAVLNGAVVSTPHYSPGLLIEKSDAYTKVYSRAGLTLMWNREDALMLELDTK
                     FRNHTCGLCGDYNGLQSYSEFLSDGVLFSPLEFGNMQKINQPDVVCEDPEEEVAPASC
                     SEHRAECERLLTAEAFADCQDLVPLEPYLRACQQDRCRCPGGDTCVCSTVAEFSRQCS
                     HAGGRPGNWRTATLCPKTCPGNLVYLESGSPCMDTCSHLEVSSLCEEHRMDGCFCPEG
                     TVYDDIGDSGCVPVSQCHCRLHGHLYTPGQEITNDCEQCVCNAGRWVCKDLPCPGTCA
                     LEGGSHITTFDGKTYTFHGDCYYVLAKGDHNDSYALLGELAPCGSTDKQTCLKTVVLL
                     ADKKKNAVVFKSDGSVLLNQLQVNLPHVTASFSVFRPSSYHIMVSMAIGVRLQVQLAP
                     VMQLFVTLDQASQGQVQGLCGNFNGLEGDDFKTASGLVEATGAGFANTWKAQSTCHDK
                     LDWLDDPCSLNIESANYAEHWCSLLKKTETPFGRCHSAVDPAEYYKRCKYDTCNCQNN
                     EDCLCAALSSYARACTAKGVMLWGWREHVCNKDVGSCPNSQVFLYNLTTCQQTCRSLS
                     EADSHCLEGFAPVDGCGCPDHTFLDEKGRCVPLAKCSCYHRGLYLEAGDVVVRQEERC
                     VCRDGRLHCRQIRLIGQSCTAPKIHMDCSNLTALATSKPRALSCQTLAAGYYHTECVS
                     GCVCPDGLMDDGRGGCVVEKECPCVHNNDLYSSGAKIKVDCNTCTCKRGRWVCTQAVC
                     HGTCSIYGSGHYITFDGKYYDFDGHCSYVAVQDYCGQNSSLGSFSIITENVPCGTTGV
                     TCSKAIKIFMGRTELKLEDKHRVVIQRDEGHHVAYTTREVGQYLVVESSTGIIVIWDK
                     RTTVFIKLAPSYKGTVCGLCGNFDHRSNNDFTTRDHMVVSSELDFGNSWKEAPTCPDV
                     STNPEPCSLNPHRRSWAEKQCSILKSSVFSICHSKVDPKPFYEACVHDSCSCDTGGDC
                     ECFCSAVASYAQECTKEGACVFWRTPDLCPIFCDYYNPPHECEWHYEPCGNRSFETCR
                     TINGIHSNISVSYLEGCYPRCPKDRPIYEEDLKKCVTADKCGCYVEDTHYPPGASVPT
                     EETCKSCVCTNSSQVVCRPEEGKILNQTQDGAFCYWEICGPNGTVEKHFNICSITTRP
                     STLTTFTTITLPTTPTSFTTTTTTTTPTSSTVLSTTPKLCCLWSDWINEDHPSSGSDD
                     GDREPFDGVCGAPEDIECRSVKDPHLSLEQHGQKVQCDVSVGFICKNEDQFGNGPFGL
                     CYDYKIRVNCCWPMDKCITTPSPPTTTPSPPPTTTTTLPPTTTPSPPTTTTTTPPPTT
                     TPSPPITTTTTPLPTTTPSPPISTTTTPPPTTTPSPPTTTPSPPTTTPSPPTTTTTTP
                     PPTTTPSPPMTTPITPPASTTTLPPTTTPSPPTTTTTTPPPTTTPSPPTTTPITPPTS
                     TTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPPTITTTTPPPTTTPSPPTTTTT
                     TPPPTTTPSPPTTTPITPPTSTTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPP
                     ITTTTTPPPTTTPSSPITTTPSPPTTTMTTPSPTTTPSSPITTTTTTPSSTTTPSPPPT
                     TMTTPSPTTTPSPPTTTMTTLPPTTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVP
                     LCNWTGWLDSGKPNFHKPGGDTELIGDVCGPGWAANISCRATMYPDVPIGQLGQTVVC
                     DVSVGLICKNEDQKPGGVIPMAFCLNYEINVQCCECVTQPTTMTTTTTENPTPPTTTP
                     ITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPT
                     PTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPI
                     TTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTP
                     TGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPIT
                     TTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPT
                     GTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITT
                     TTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTG
                     TQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTT
                     TTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGT
                     QTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTT
                     TVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQ
                     TPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTT
                     VTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQT
                     PTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTV
                     TPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTP
                     TTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVT
                     PTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPT
                     TTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTP
                     TPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTT
                     TPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPT
                     PTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTT
                     PITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTP
                     TPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTP
                     ITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPT
                     PTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPI
                     TTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTP
                     TGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPIT
                     TTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPT
                     GTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITT
                     TTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTG
                     TQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTT
                     TTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGT
                     QTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTT
                     TVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQ
                     TPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTT
                     VTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQT
                     PTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTV
                     TPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTP
                     TTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVT
                     PTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQGPPTHTSTAPIAELTTSNPPPESST
                     PQTSRSTSSPLTESTTLLSTLPPAIEMTSTAPPSTPTAPTTTSGGHTLSPPPSTTTSP
                     PGTPTRGTTTGSSSAPTPSTVQTTTTSAWTPTPTPLSTPSIIRTTGLRPYPSSVLICC
                     VLNDTYYAPGEEVYNGTYGDTCYFVNCSLSCTLEFYNWSCPSTPSPTPTPSKSTPTPS
                     KPSSTPSKPTPGTKPPECPDFDPPRQENETWWLCDCFMATCKYNNTVEIVKVECEPPP
                     MPTCSNGLQPVRVEDPDGCCWHWECDCYCTGWGDPHYVTFDGLYYSYQGNCTYVLVEE
                     ISPSVDNFGVYIDNYHCDPNDKVSCPRTLIVRHETQEVLIKTVHMMPMQVQVQVNRQA
                     VALPYKKYGLEVYQSGINYVVDIPELGVLVSYNGLSFSVRLPYHRFGNNTKGQCGTCT
                     NTTSDDCILPSGEIVSNCEAAADQWLVNDPSKPHCPHSSSTTKRPAVTVPGGGKTTPH
                     KDCTPSPLCQLIKDSLFAQCHALVPPQHYYDACVFDSCFMPGSSLECASLQAYAALCA
                     QQNICLDWRNHTHGACLVECPSHREYQACGPAEEPTCKSSSSQQNNTVLVEGCFCPEG
                     TMNYAPGFDVCVKTCGCVGPDNVPREFGEHFEFDCKNCVCLEGGSGIICQPKRCSQKP
                     VTHCVEDGTYLATEVNPADTCCNITVCKCNTSLCKEKPSVCPLGFEVKSKMVPGRCCP
                     FYWCESKGVCVHGNAEYQPGSPVYSSKCQDCVCTDKVDNNTLLNVIACTHVPCNTSCS
                     PGFELMEAPGECCKKCEQTHCIIKRPDNQHVILKPGDFKSDPKNNCTFFSCVKIHNQL
                     ISSVSNITCPNFDASICIPGSITFMPNGCCKTCTPRNETRVPCSTVPVTTEVSYAGCT
                     KTVLMNHCSGSCGTFVMYSAKAQALDHSCSCCKEEKTSQREVVLSCPNGGSLTHTYTH
                     IESCQCQDTVCGLPTGTSRRARRSPRHLGSG"
     misc_feature    133. .1194
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     misc_feature    1195. .2295
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     misc_feature    2602. .3687
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     repeat_region   4228. .5268
                     /note="repetitive Thr/Ser/Pro-rich subdomain"
     misc_feature    13465. .14634
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
BASE COUNT     4107 a   6921 c   2862 g   1830 t
ORIGIN

  Query Match             7.9%;  Score 193;  DB 11;  Length 15720;
  Best Local Similarity   46.1%;  Pred. No. 3.7e-27;
  Matches  874;  Conservative    0;  Mismatches 1000;  Indels   21;  Gaps    6;

Qy     256 CGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCTGCC 315
           | | | | || ||  || ||||    ||||     | | |   | ||     ||      |
Db    5628 CAACTACGAGATCAACGTTCAGTGCTGTGAGTGTGTCACCCAACCCACCACCATGACAAC 5687

Qy     316 AGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCACCCA 375
                |  |||    |  |||| | | |   |  | ||| |      || | || || ||
Db    5688 CACCACCACAGAGAACCCAACTCCGCCAACCACGACACCCATCACCACCACCACTACGGT 5747

Qy     376 GGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAACCAC 435
           | |||||||| ||||   |     | | |     ||  | |||  | | || | ||| |||||
Db    5748 GACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCAC 5807

Qy     436 CCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCAGGC 495
           |    |     ||| | ||||    | || |||   ||| ||| ||||||| || || |
Db    5808 CACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCAT 5867

Qy     496 CCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCACCCA 555
           | | || || || ||    | |||||||| ||||  |     | | |     ||  | |||
Db    5868 CACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCAC 5927

Qy     556 GGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGCCCC 615
```

```
Db   5928 GACACCCATCACCACC---ACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACA 5984

Qy    616 AACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTACCCC 675
Db   5985 GACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCAC 6044

Qy    676 GGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGCACC 735
Db   6045 CGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCC 6104

Qy    736 CACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAG---- 792
Db   6105 AACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGAC 6164

Qy    792 -CTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAG 850
Db   6165 CCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCAC 6224

Qy    851 TGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTAC 910
Db   6225 TACGGTGACCCCAACCCCAACACCCACCGGCACACAGAC---CCCAACCACGACACCCAT 6281

Qy    911 GTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCC 970
Db   6282 CACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCAC 6341

Qy    971 CCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGA 1030
Db   6342 GACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGAC 6401

Qy   1031 TGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTAT 1090
Db   6402 CCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGG 6461

Qy   1091 TACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGT 1150
Db   6462 CACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAAC 6521

Qy   1151 TTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAG 1210
Db   6522 ACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCC 6581

Qy   1211 AGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCC 1270
Db   6582 AACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTAC 6641

Qy   1271 ATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTC 1330
Db   6642 GGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCAC 6701

Qy   1331 CCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGA 1390
Db   6702 CACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACC--CCAACCACGACA 6759

Qy   1391 CCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACC 1450
Db   6760 CCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCA 6819

Qy   1451 GCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCG 1510
Db   6820 ACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACA 6879

Qy   1511 ACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCC 1570
Db   6880 CAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCC 6939

Qy   1571 TCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCC------AGTCCCACCT 1624
Db   6940 ACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACC 6999

Qy   1625 ACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCT 1684
```

```
Db   7000 CCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTG 7059

Qy   1685 GCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1744
Db   7060 ACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACC 7119

Qy   1745 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1804
Db   7120 ACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATC 7179

Qy   1805 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1864
Db   7180 ACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACG 7239

Qy   1865 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1924
Db   7240 ACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACC 7299

Qy   1925 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTC 1984
Db   7300 CCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGC 7359

Qy   1985 ACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCAAC 2044
Db   7360 ACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACA 7419

Qy   2045 ACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCT--GAAGATGATG 2102
Db   7420 CCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCA 7479

Qy   2103 ACATCGACCCCCCTCTCCCCAACGACCCCATCAAC 2137
Db   7480 ACCCCAACACCCACCGGCACACAGACCCCAACCAC 7514
```

```
RESULT   4
HUMMUC2X
LOCUS       HUMMUC2X    15720 bp    mRNA            PRI       09-JAN-1998
DEFINITION  Homo sapiens intestinal mucin (MUC2) mRNA, complete cds.
ACCESSION   L21998
NID         g454153
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 15720)
  AUTHORS   Gum,J.R.Jr. ., Hicks,J.W., Toribara,N.W., Kim,Y.S. and Siddiki,B.
  TITLE     Molecular cloning of human intestinal mucin (MUC2) cDNA:
            Identification of the amino terminal and overall sequence
            similarity to pre-pro-von Willebrand factor
  JOURNAL   J. Biol. Chem. 269, 2440-2446 (1994)
  MEDLINE   94132002
REFERENCE   2  (bases 1 to 15720)
  AUTHORS   Eckhardt,A.E., Timpte,C.S., DeLuca,A.W. and Hill,R.L.
  TITLE     The complete cDNA sequence and structural polymorphism of the
            polypeptide chain of porcine submaxillary mucin
  JOURNAL   J. Biol. Chem. 272 (52), 33204-33210 (1997)
  MEDLINE   98070526
FEATURES             Location/Qualifiers
     source          1. .15720
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="intestine"
     CDS             28. .15567
                     /codon_start=1
                     /product="mucin"
                     /db_xref="PID:g454154"
                     /translation="MGLPLARLAAVCLALSLAGGSELQTEGRTRYHGRNVCSTWGNFH
                     YKTFDGDVFRFPGLCDYNFASDCRGSYKEFAVHLKRGPGQAEAPAGVESILLTIKDDT
                     IYLTRHLAVLNGAVVSTPHYSPGLLIEKSDAYTKVYSRAGLTLMWNREDALMLELDTK
                     FRNHTCGLCGDYNGLQSYSEFLSDGVLFSPLEFGNMQKINQPDVVCEDPEEEVAPASC
                     SEHRAECERLLTAEAFADCQDLVPLEPYLRACQQDRCRCPGGDTCVCSTVAEFSRQCS
```

```
HAGGRPGNWRTATLCPKTCPGNLVYLESGSPCMDTCSHLEVSSLCEEHRMDGCFCPEG
TVYDDIGDSGCVPVSQCHCRLHGHLYTPGQEITNDCEQCVCNAGRWVCKDLPCPGTCA
LEGGSHITTFDGKTYTFHGDCYYVLAKGDHNDSYALLGELAPCGSTDKQTCLKTVVLL
ADKKKNAVVFKSDGSVLLNQLQVNLPHVTASFSVFRPSSYHIMVSMAIGVRLQVQLAP
VMQLFVTLDQASQGQVQGLCGNFNGLEGDDFKTASGLVEATGAGFANTWKAQSTCHDK
LDWLDDPCSLNIESANYAEHWCSLLKKTETPFGRCHSAVDPAEYYKRCKYDTCNCQNN
EDCLCAALSSYARACTAKGVMLWGWREHVCNKDVGSCPNSQVFLYNLTTCQQTCRSLS
EADSHCLEGFAPVDGCGCPDHTFLDEKGRCVPLAKCSCYHRGLYLEAGDVVVRQEERC
VCRDGRLHCRQIRLIGQSCTAPKIHMDCSNLTALATSKPRALSCQTLAAGYYHTECVS
GCVCPDGLMDDGRGGCVVEKECPCVHNNDLYSSGAKIKVDCNTCTCKRGRWVCTQAVC
HGTCSIYGSGHYITFDGKYYDFDGHCSYVAVQDYCGQNSSLGSFSIITENVPCGTTGV
TCSKAIKIFMGRTELKLEDKHRVVIQRDEGHHVAYTTREVGQYLVVESSTGIIVIWDK
RTTVFIKLAPSYKGTVCGLCGNFDHRSNNDFTTRDHMVVSSELDFGNSWKEAPTCPDV
STNPEPCSLNPHRRSWAEKQCSILKSSVFSICHSKVDPKPFYEACVHDSCSCDTGGDC
ECFCSAVASYAQECTKEGACVFWRTPDLCPIFCDYYNPPHECEWHYEPCGNRSFETCR
TINGIHSNISVSYLEGCYPRCPKDRPIYEEDLKKCVTADKCGCYVEDTHYPPGASVPT
EETCKSCVCTNSSQVVCRPEEGKILNQTQDGAFCYWEICGPNGTVEKHFNICSITTRP
STLTTFTTITLPTTPTSFTTTTTTTTPTSSTVLSTTPKLCCLWSDWINEDHPSSGSDD
GDREPFDGVCGAPEDIECRSVKDPHLSLEQHGQKVQCDVSVGFICKNEDQFGNGPFGL
CYDYKIRVNCCWPMDKCITTPSPPTTTPSPPPTTTTTLPPTTTPSPPTTTTTTPPPTT
TPSPPITTTTTPLPTTTPSPPISTTTTPPPTTTPSPPTTTPSPPTTTPSPPTTTTTTP
PPTTTPSPPMTTPITPPASTTTLPPTTTPSPPTTTTTPPPTTTPSPPTTTPITPPTS
TTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPPTITTTTPPPTTTPSPPTTTTT
TPPPTTTPSPPTTTPITPPTSTTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPP
ITTTTTPPPTTTPSSPITTTPSPPTTTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPT
TMTTPSPTTTPSPPTTTMTTLPPTTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVP
LCNWTGWLDSGKPNFHKPGGDTELIGDVCGPGWAANISCRATMYPDVPIGQLGQTVVC
DVSVGLICKNEDQKPGGVIPMAFCLNYEINVQCCECVTQPTTMTTTTTENPTPPTTTP
ITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT
PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI
TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP
TGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPIT
TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPT
GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITT
TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTG
TQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTT
TTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGT
QTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTT
TVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ
TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTT
VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT
PTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTV
TPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTP
TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT
PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT
TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTP
TPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTT
TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPT
PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTT
PITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP
TPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTP
ITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT
PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI
TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP
TGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPIT
TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPG
TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTG
TQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTT
TTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGT
QTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTT
TVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ
TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTT
VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT
PTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTV
TPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTP
TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT
PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPT
TTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTP
TPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTT
TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPT
PTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTT
PITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTP
TPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTP
ITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPT
PTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI
TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP
TGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPIT
TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPG
TQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTT
TTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGT
QTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTT
TVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ
TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTT
VTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT
PTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTV
TPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTP
TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT
PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTGPPTHTSTAPIAELTTSNPPPESST
PQTSRSTSSPLTESTTLLSTLPPAIEMTSTAPPSTPTAPTTTSGGHTLSPPPSTTTSP
PGTPTRGTTTGSSSAPTPSTVQTTTTSAWTPTPTPLSTPSIIRTTGLRPYPSSVLICC
VLNDTYYAPGEEVYNGTYGDTCYFVNCSLSCTLEFYNWSCPSTPSPTPTPSKSTPTPS
KPSSTPSKPTPGTKPPECPDFDPPRQENETWWLCDCFMATCKYNNTVEIVKVECEPPP
MPTCSNGLQPVRVEDPDGCCWHWECDCYCTGWGDPHYVTFDGLYYSYQGNCTYVLVEE
```

```
                ISPSVDNFGVYIDNYHCDPNDKVSCPRTLIVRHETQEVLIKTVHMMPMQVQVQVNRQA
                VALPYKKYGLEVYQSGINYVVDIPELGVLVSYNGLSFSVRLPYHRFGNNTKGQCGTCT
                NTTSDDCILPSGEIVSNCEAAADQWLVNDPSKPHCPHSSSTTKRPAVTVPGGGKTTPH
                KDCTPSPLCQLIKDSLFAQCHALVPPQHYYDACVFDSCFMPGSSLECASLQAYAALCA
                QQNICLDWRNHTHGACLVECPSHREYQACGPAEEPTCKSSSSQQNNTVLVEGCFCPEG
                TMNYAPGFDVCVKTCGCVGPDNVPREFGEHFEFDCKNCVCLEGGSGIICQPKRCSQKP
                VTHCVEDGTYLATEVNPADTCCNITVCKCNTSLCKEKPSVCPLGFEVKSKMVPGRCCP
                FYWCESKGVCVHGNAEYQPGSPVYSSKCQDCVCTDKVDNNTLLNVIACTHVPCNTSCS
                PGFELMEAPGECCKKCEQTHCIIKRPDNQHVILKPGDFKSDPKNNCTFFSCVKIHNQL
                ISSVSNITCPNFDASICIPGSITFMPNGCCKTCTPRNETRVPCSTVPVTTEVSYAGCT
                KTVLMNHCSGSCGTFVMYSAKAQALDHSCSCCKEEKTSQREVVLSCPNGGSLTHTYTH
                IESCQCQDTVCGLPTGTSRRARRSPRHLGSG"
     misc_feature    133. .1194
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     misc_feature    1195. .2295
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     misc_feature    2602. .3687
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
     repeat_region   4228. .5268
                     /note="repetitive Thr/Ser/Pro-rich subdomain"
     misc_feature    13465. .14634
                     /note="Pre-pro-von Willebrand factor D domain-like
                     subdomain"
BASE COUNT     4107 a   6921 c   2862 g   1830 t
ORIGIN
```

```
  Query Match             7.9%;  Score 193;  DB 41;  Length 15720;
  Best Local Similarity   46.1%;  Pred. No. 3.7e-27;
  Matches  874;  Conservative    0;  Mismatches 1000;  Indels   21;  Gaps    6;

Qy     256 CGAATTCAAGTTCTCCGCTCAGACTTGTGTTCACGCCGCTTTAGCCGGATGCACCCTGCC 315
           | | | | || ||  || ||||   ||||     | | |   | ||     ||      |
Db    5628 CAACTACGAGATCAACGTTCAGTGCTGTGAGTGTGTCACCCAACCCACCACCATGACAAC 5687

Qy     316 AGGACCTCCAGCTGAGACAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAACCACCCA 375
             |  |||   |  |||| | | |  |  | ||| |     || | || || ||
Db    5688 CACCACCACAGAGAACCCAACTCCGCCAACCACGACACCCATCACCACCACCACTACGGT 5747

Qy     376 GGCCCCAACCACAACTACTCAGGCCCCTACTACAACCACCCAGGCCCCAACCACAACCAC 435
           | |||||||| |||| |    | | |    ||  | |||  | | || | ||| |||||
Db    5748 GACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCAC 5807

Qy     436 CCAGGCCCCAACCACCACCCAGGCCCCAACCACCACCCAGGCCCCAACTACCACTCAGGC 495
           |    |   ||| | ||||   | || |||  ||| ||| |||||| || || |
Db    5808 CACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCAT 5867

Qy     496 CCCTACTACTACCACTCAGGCCCCAACCACAACCACTCAGGCCCCTACCACAACCACCCA 555
           | | || || || ||   | |||||||| |||| |    | | |    ||  | |||
Db    5868 CACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCAC 5927

Qy     556 GGCCCCAACCACCACCCAGGCCCCAACTACCACCCAGGCCCCAACTACCACTCAGGCCCC 615
           | | || | |||||||    || | ||   ||||  |||||||  ||||     | |
Db    5928 GACACCCATCACCACC---ACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACA 5984

Qy     616 AACTACAATCACCCAGGCTGCAACTACCCCGGCCGCAACTACCCCGGCCGCAACTACCCC 675
           ||  ||| |||     |    || ||| | |  |    |||||  || ||||  || |
Db    5985 GACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCAC 6044

Qy     676 GGCCGCAACTACCCCTGCCGCGACAACCCCCGCTGCAACTACCCCAGGTGTTCCTGCACC 735
           | | ||   |||||  || ||||| ||  | |  | || ||  | |     ||  | ||
Db    6045 CGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCC 6104

Qy     736 CACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACTGTTGCCCAATGGTTGCCCAG---- 792
           ||  |  ||        | |||| | |     ||   |  |||          | |
Db    6105 AACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGAC 6164

Qy     792 -CTGACTTCGACATCCACTTGTTGATTCCCCACGACAAGTACTGCAACCTCTTCTACCAG 850
            |  ||  | ||| |||| |   |    || | || |  ||  | | | |  | ||||
Db    6165 CCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCAC 6224
```

```
Qy    851 TGCTCCAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTAC 910
Db   6225 TACGGTGACCCCAACCCCAACACCCACCGGCACACAGAC---CCCAACCACGACACCCAT 6281

Qy    911 GTCCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGACGGCGAAATCAGCCCCGCACCC 970
Db   6282 CACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCAC 6341

Qy    971 CCAGTCACAGAAGGCAACGAAGACGAAGACATTGACATCGGAGACCTCCTCGACAATGGA 1030
Db   6342 GACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGAC 6401

Qy   1031 TGCCCAGCTAACTTCGAAATCGACTGGCTCTTGCCCCACGGAAACCGTTGCGACAAGTAT 1090
Db   6402 CCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGG 6461

Qy   1091 TACCAGTGCGTCCACGGTAACTTGGTAGAGAGGCGTTGTGGAGCCGGCACCCACTTCAGT 1150
Db   6462 CACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAAC 6521

Qy   1151 TTTGAACTTCAGCAATGTGACCACATCGAGCTCGTTGGCTGCACCCTCCCCGGCGGCGAG 1210
Db   6522 ACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCC 6581

Qy   1211 AGCGAAGAAGTTGACGTCGACGAGGATGCCTGCACCGGCTGGTACTGCCCCACGGAACCC 1270
Db   6582 AACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTAC 6641

Qy   1271 ATTGAATGGGAGCCCCTCCCCAACGGCTGCCCTGCCGACTTCAGCATCGACCACCTCCTC 1330
Db   6642 GGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCAC 6701

Qy   1331 CCCCACGAGAGCGACTGCGGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGA 1390
Db   6702 CACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACC--CCAACCACGACA 6759

Qy   1391 CCTTGCCCTGGAAACCTGCACTTCAGTCCTGCCACACAGTCCTGTGAGTCTCCTGTGACC 1450
Db   6760 CCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCA 6819

Qy   1451 GCTGGTTGCCAAGTTTTCGAGTGCGATTCTGACAACCAGTGCACATCGACTGCTGCCCCG 1510
Db   6820 ACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACA 6879

Qy   1511 ACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAACTGCCGCACCC 1570
Db   6880 CAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCC 6939

Qy   1571 TCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCC------AGTCCCACCT 1624
Db   6940 ACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACC 6999

Qy   1625 ACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTACTGCT 1684
Db   7000 CCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTG 7059

Qy   1685 GCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAGCCCCT 1744
Db   7060 ACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACC 7119

Qy   1745 ACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTGCCCCCACC 1804
Db   7120 ACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACGACACCCATC 7179

Qy   1805 GCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAATCCCAACT 1864
Db   7180 ACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCACG 7239

Qy   1865 ACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCAACACCACA 1924
Db   7240 ACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGCACACAGACC 7299

Qy   1925 GTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACACCACAGTC 1984
Db   7300 CCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACACCCACCGGC 7359

Qy   1985 ACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATGCACCCAAC 2044
Db   7360 ACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCAACCCCAACA 7419

Qy   2045 ACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCT--GAAGATGATG 2102
Db   7420 CCCACCGGCACACAGACCCCAACCACGACACCCATCACCACCACCACTACGGTGACCCCA 7479

Qy   2103 ACATCGACCCCCCTCTCCCCAACGACCCCATCAAC 2137
Db   7480 ACCCCAACACCCACCGGCACACAGACCCCAACCAC 7514
```

RESULT 5
AC002042/c
LOCUS AC002042 185994 bp DNA HTG 22-APR-1998
DEFINITION *** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 16p11.2 BAC clone CIT987SK-A-180G2; HTGS phase 1, 5 unordered pieces.
ACCESSION AC002042
NID g3075381
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 185994)
AUTHORS Adams,M.D., Loftus,B.J., Zhou,L., Phillips,C., Brandon,R.C., Fuhrmann,J., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
TITLE Human chromosome 16p13 BAC clone CIT987SK-A-180G2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185994)
AUTHORS Adams,M.D. and Loftus,B.J.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Apr 22, 1998 this sequence version replaced gi:3068565.
***
*** WARNING: Phase 1 High Throughput Genome Sequence ***
***
* This sequence is unfinished. It consists of 5 contigs for
* which the order is not known; their order in this record is
* arbitrary. In some cases, the exact lengths of the gaps
* between the contigs are also unknown; these gaps are presented
* as runs of N as a convenience only. When sequencing is complete,
* the sequence data presented in this record will be replaced
*by a single finished sequence with the same accession number.
* 1 2198: contig of 2198 bp in length
* 2199 2248: gap of unknown length
* 2249 46288: contig of 44040 bp in length
* 46289 46338: gap of unknown length
* 46339 110919: contig of 64581 bp in length
* 110920 110969: gap of unknown length
* 110970 145225: contig of 34256 bp in length
* 145226 145275: gap of unknown length
* 145276 185994: contig of 40719 bp in length.
FEATURES Location/Qualifiers
source 1. .185994
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="A-180G2"
BASE COUNT 51275 a 43334 c 43047 g 48137 t 201 others
ORIGIN

Query Match 6.2%; Score 151.8; DB 18; Length 185994;
Best Local Similarity 53.7%; Pred. No. 1.6e-19;
Matches 337; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

```
Qy   1497 CGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCC 1556
```

```
Db 111975 CTACTCCTCCTGCCACAGCTACTGCAACTGCTGCTACTCCTCCTGCCACAGCTACTGCAA 111916

Qy   1557 CAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAG 1616
Db 111915 CTGCTGCTGCTGCTACTCCTCCTGCCACAGCTACTGCAACTACTGATGCTGCTACTCCTG 111856

Qy   1617 TCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTA 1676
Db 111855 CCACAGCTACTGCAACTGCTGATGCTGCTACTCCTCCTGCCACAGCTACTGCAACTGCTG 111796

Qy   1677 CTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCG 1736
Db 111795 ATGCTGCTACTCCTCCTGCCACAGCTACTGCAACTGCTGATGCTGCTACTCCTCCTGCCA 111736

Qy   1737 CAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCTG 1796
Db 111735 CAGCTACTGCAACTGCTGCTGCTGCTACTCCTCCTGCCACAGCTACTGCAACTGCCGATG 111676

Qy   1797 CCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAGAAA 1856
Db 111675 CTGCTACTCCTGCCACAGCTACTGCAACTGCTGATGCTGCTACTCCTCCTGCCACAGCTA 111616

Qy   1857 TCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCCCCA 1916
Db 111615 CTGCAACTGCTGATGCTGCTACTCCTCCTGCCACAGCTACTGCAACTGCTGATGCTGCTA 111556

Qy   1917 ACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCAACA 1976
Db 111555 CTCCTCCTGCCACAGCTACTGCAACTGCTGATGCTGCTACTCCT---CCTGCCACAGCTA 111499

Qy   1977 CCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCACATG 2036
Db 111498 CTGCAACTGCTGATGCTGCTACTCCTCCTGCCACAGCTACTGCAACTACTGATGCTGCTA 111439

Qy   2037 CACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTGAAG 2096
Db 111438 CTCCTGCCACAGCTACTGCAACTGCTGATGCTGCTACTCCTCCTGCCACAGCTACTGCAA 111379

Qy   2097 ATGATGACATCGACCCCCCTCTCCCCA 2123
Db 111378 CTGCTGATGCTGCTACTCCTCCTGCCA 111352
```

```
RESULT   6
I25041/c
LOCUS       I25041      10596 bp    DNA             PAT       21-AUG-1996
DEFINITION  Sequence 15 from patent US 5547856.
ACCESSION   I25041
NID         g1604911
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10596)
  AUTHORS   Godowski,P.J., Lokker,N.A. and Mark,M.R.
  TITLE     Hepatocyte growth factor variants
  JOURNAL   Patent: US 5547856-A 15 20-AUG-1996;
FEATURES             Location/Qualifiers
     source          1. .10596
                     /organism="unknown"
BASE COUNT     2627 a   2571 c   3023 g   2375 t
ORIGIN
```

```
  Query Match             6.2%;  Score 151.6;  DB 6;  Length 10596;
  Best Local Similarity   53.2%;  Pred. No. 2.2e-19;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
Db   2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
Db   2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
Db   2747 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 2688

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
Db   2687 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
Db   2627 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
Db   2567 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
Db   2507 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
Db   2447 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 2388

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
Db   2387 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
Db   2327 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216
```

```
RESULT   7
I30503/c
LOCUS       I30503      10596 bp    DNA             PAT       13-DEC-1996
DEFINITION  Sequence 15 from patent US 5580963.
ACCESSION   I30503
NID         g1821294
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 10596)
  AUTHORS   Godowski,P.J., Lokker,N.A. and Mark,M.R.
  TITLE     Single-chain hepatocyte growth factor variants
  JOURNAL   Patent: US 5580963-A 15 03-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .10596
                     /organism="unknown"
BASE COUNT     2627 a   2571 c   3023 g   2375 t
ORIGIN
```

```
  Query Match             6.2%;  Score 151.6;  DB 6;  Length 10596;
  Best Local Similarity   53.2%;  Pred. No. 2.2e-19;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
Db   2867 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 2808

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
Db   2807 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 2748
```

```
Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
Db   2747 CTGCCCCTCCTGCCCCCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCTCCTGCCCCCTCCTG 2688

Qy   1680 CTGCTGCCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
Db   2687 CTCCTGCCCCCTCCTGCTCCTGCCCCCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTC 2628

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
Db   2627 CTCCTGCTCCTGCCCCCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCTCCTG 2568

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCCTACTACTGCTGTCCCAG 1853
Db   2567 CCCCTCCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTCCTGCTCCTGCCCCCTC 2508

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
Db   2507 CTGCCCCCTCCTCCTGCTCCTGCCCCCTCCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCCC 2448

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCCCCCA 1973
Db   2447 CTCCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTCCTGCTCCTGCCCCCTCCTCCTGCTCCTG 2388

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
Db   2387 CCCCTCCTGCCCCCTCCTGCCCCCTCCTCCTGCTCCTGCCCCCTCCTCCTGCTCCTGCCCCCTC 2328

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
Db   2327 CTGCCCCTCCTGCCCCCTCCTGCCCCCTCCTCCTGCTCCTGCCCCCTCCTCCTGCTCCTGCCC 2268

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
Db   2267 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 2216
```

RESULT 8
U02454/c
LOCUS U02454 5452 bp DNA circular SYN 29-MAR-1996
DEFINITION Cloning vector pCMVEBNA, complete sequence.
ACCESSION U02454
NID g413820
KEYWORDS .
SOURCE Cloning vector pCMVEBNA.
ORGANISM Cloning vector pCMVEBNA
artificial sequence; cloning vectors.
REFERENCE 1 (bases 1 to 5452)
AUTHORS Kitts,P.A.
TITLE CLONTECH Vectors On Disc version 1.3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5452)
AUTHORS Swirski,R.A., Van Den Berg,D., Murphy,A.J., Lambert,C.M., Friedberg,E.C. and Schimke,R.T.
TITLE Improvements in the Epstein-Barr-based shuttle vector system for direct cloning in human tissue culture cells
JOURNAL Methods: A Companion to Methods in Enzymology 4, 133-142 (1992)
REFERENCE 3 (bases 1 to 5452)
AUTHORS Kitts,P.A.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT This vector can be obtained from CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303, USA. To place an order call (415) 424-8222 or (800) 662-2566, extension 1. International customers, please contact your local distributor. For technical information, call (415) 424- 8222 or (800) 662-2566, extension 3. This sequence has been compiled from information in the sequence databases, published literature and other sources, together with partial sequences obtained by CLONTECH; this vector has not been completely sequenced. If you suspect there is an error in this sequence, please contact CLONTECH's Technical Service Department at (415) 424-8222 or (800) 662-2566, extension 3 or E-mail TECH@CLONTECH.COM.
FEATURES Location/Qualifiers
source 1. .5452
/organism="Cloning vector pCMVEBNA"
/db_xref="taxon:31797"
BASE COUNT 1362 a 1246 c 1736 g 1108 t
ORIGIN

Query Match 6.2%; Score 151.6; DB 15; Length 5452;
Best Local Similarity 53.2%; Pred. No. 2.4e-19;
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

```
Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
Db   1701 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 1642

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
Db   1641 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 1582

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
Db   1581 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 1522

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
Db   1521 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 1462

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
Db   1461 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 1402

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCCTACTACTGCTGTCCCAG 1853
Db   1401 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 1342

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
Db   1341 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 1282

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
Db   1281 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 1222

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
Db   1221 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 1162

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
Db   1161 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 1102

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
Db   1101 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 1050
```

RESULT 9
U02455/c
LOCUS U02455 10850 bp DNA circular SYN 29-MAR-1996
DEFINITION Cloning vector rpDR2, complete sequence.
ACCESSION U02455
NID g413821
KEYWORDS .
SOURCE Cloning vector rpDR2.
ORGANISM Cloning vector rpDR2
artificial sequence; cloning vectors.
REFERENCE 1 (bases 1 to 10850)
AUTHORS Kitts,P.A.
TITLE CLONTECH Vectors On Disc version 1.3
JOURNAL Unpublished

```
REFERENCE   2  (bases 1 to 10850)
  AUTHORS   Murphy,A.J., Kung,A.L., Swirski,R.A. and Schimke,R.T.
  TITLE     cDNA expression cloning in human cells using the plambdaDR2
            episomal vector system
  JOURNAL   Methods: A Companion to Methods in Enzymology 4, 111-131 (1992)
REFERENCE   3  (bases 1 to 10850)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT     Lambda DR2 can be obtained from CLONTECH Laboratories, Inc., 1020
            East Meadow Circle, Palo Alto, CA 94303, USA. To place an order
            call (415) 424-8222 or (800) 662-2566, extension 1. International
            customers, please contact your local distributor.  For technical
            information, call (415) 424- 8222 or (800) 662-2566, extension 3.
            This sequence was compiled by Andrew Murphy and revised at
            CLONTECH. If you suspect there is an error in this sequence, please
            contact CLONTECH's Technical Service Department at (415) 424-8222
            or (800) 662-2566, extension 3 or E-mail TECH@CLONTECH.COM.
FEATURES             Location/Qualifiers
     source          1. .10850
                     /organism="Cloning vector rpDR2"
                     /note="plasmid released from lambda DR2"
                     /db_xref="taxon:31858"
BASE COUNT     2723 a   2615 c   3053 g   2459 t
ORIGIN
```

```
  Query Match             6.2%;  Score 151.6;  DB 15;  Length 10850;
  Best Local Similarity   53.2%;  Pred. No. 2.2e-19;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || |  |  |  |||||||  | ||| |  |  |  |||||||
Db   7347 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 7288

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db   7287 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 7228

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db   7227 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 7168

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || |||||||||    | ||  || ||    |     | ||| ||||  ||||  || |
Db   7167 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 7108

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||       |     ||| | |  | || ||| | |||
Db   7107 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 7048

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
             | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db   7047 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 6988

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
             |||  || |||    |  | || ||| |||||  |  || || |  ||  | |||||||
Db   6987 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 6928

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |     |  |      |     ||  || || || |  || || || |   | |||||  ||
Db   6927 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 6868

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
            | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db   6867 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 6808

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
            ||| ||     || |  |||||||     || |  |   | | ||  || |  |  |
Db   6807 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 6748

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
```

```
             || |     |   ||| |||   |    |  ||  |    || || ||
Db   6747 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 6696


RESULT  10
XXU02428/c
LOCUS       XXU02428    10737 bp    DNA   circular  SYN       29-JAN-1997
DEFINITION  Cloning vector pDR2, complete sequence.
ACCESSION   U02428
NID         g413794
KEYWORDS    .
SOURCE      Cloning vector pDR2.
  ORGANISM  Cloning vector pDR2
            artificial sequence; cloning vectors.
REFERENCE   1  (bases 1 to 10737)
  AUTHORS   Murphy,A.J., Kung,A.L., Swirski,R.A. and Schimke,R.T.
  TITLE     cDNA expression cloning in human cells using the plambdaDR2
            episomal vector system
  JOURNAL   Methods: A Companion to Methods in Enzymology 4, 111-131 (1992)
REFERENCE   2  (bases 1 to 10737)
  AUTHORS   Kitts,P.A.
  TITLE     CLONTECH Vectors On Disc version 1.3
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 10737)
  AUTHORS   Kitts,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,
            1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT     This vector can be obtained from CLONTECH Laboratories, Inc., 1020
            East Meadow Circle, Palo Alto, CA 94303, USA. To place an order
            call (415) 424-8222 or (800) 662-2566, extension 1. International
            customers, please contact your local distributor.  For technical
            information, call (415) 424- 8222 or (800) 662-2566, extension 3.
            This sequence has been compiled from information in the sequence
            databases, published literature and other sources, together with
            partial sequences obtained by CLONTECH; this vector has not been
            completely sequenced. If you suspect there is an error in this
            sequence, please contact CLONTECH's Technical Service Department at
            (415) 424-8222 or (800) 662-2566, extension 3 or E-mail
            TECH@CLONTECH.COM.'.
FEATURES             Location/Qualifiers
     source          1. .10737
                     /organism="Cloning vector pDR2"
                     /db_xref="taxon:31805"
BASE COUNT     2686 a   2596 c   3033 g   2422 t
ORIGIN
```

```
  Query Match             6.2%;  Score 151.6;  DB 15;  Length 10737;
  Best Local Similarity   53.2%;  Pred. No. 2.2e-19;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || |  |  |  |||||||  | ||| |  |  |  |||||||
Db   7234 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 7175

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  ||||  | || |  |
Db   7174 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 7115

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db   7114 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 7055

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || |||||||||    | ||  || ||    |     | ||| ||||  ||||  || |
Db   7054 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 6995

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||       |     ||| | |  | || ||| | |||
Db   6994 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 6935

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
```

```
Db    6934 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 6875

Qy    1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913

Db    6874 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 6815

Qy    1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973

Db    6814 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 6755

Qy    1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033

Db    6754 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 6695

Qy    2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093

Db    6694 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 6635

Qy    2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145

Db    6634 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 6583


RESULT  11
EBV/c
LOCUS       EBV        172281 bp    DNA   circular  VRL       17-AUG-1998
DEFINITION  Epstein-Barr virus (EBV) genome. The complete sequence [1-10] was
            determined from DNA from B95-8 cells cloned by Arrand et al [11].
            B95-8 is a productive marmoset lymphoblastoid cell line
            immortalized with human EBV from a mononucleosis patient.
ACCESSION   V01555 J02070 K01729 K01730 V01554 X00498 X00499 X00784
NID         g59074
KEYWORDS    DNA polymerase; EBNA; genome; ribonucleotide reductase; tandem
            repeat; terminal repeat.
SOURCE      Epstein-Barr virus.
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 172281)
  AUTHORS   Arrand,J.R., Rymo,L., Walsh,J.E., Bjorck,E., Lindahl,T. and
            Griffin,B.E.
  TITLE     Molecular cloning of the complete Epstein-Barr virus genome as a
            set of overlapping restriction endonuclease fragments
  JOURNAL   Nucleic Acids Res. 9 (13), 2999-3014 (1981)
  MEDLINE   82014887
REFERENCE   2  (bases 1 to 172281)
  AUTHORS   Kozak,M.
  TITLE     Possible role of flanking nucleotides in recognition of the AUG
            initiator codon by eukaryotic ribosomes
  JOURNAL   Nucleic Acids Res. 9 (20), 5233-5262 (1981)
  MEDLINE   82059504
REFERENCE   3  (bases 1 to 172281)
  AUTHORS   Deininger,P.L., Bankier,A., Farrell,P., Baer,R. and Barrell,B.
  TITLE     Sequence analysis and in vitro transcription of portions of the
            Epstein-Barr virus genome
  JOURNAL   J. Cell. Biochem. 19 (3), 267-274 (1982)
  MEDLINE   83109311
REFERENCE   4  (bases 1 to 172281)
  AUTHORS   Farrell,P.J., Deininger,P.L., Bankier,A. and Barrell,B.
  TITLE     Homologous upstream sequences near Epstein-Barr virus promoters
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 80 (6), 1565-1569 (1983)
  MEDLINE   83169725
REFERENCE   5  (bases 45644 to 52450)
  AUTHORS   Jeang,K.T. and Hayward,S.D.
  TITLE     Organization of the Epstein-Barr virus DNA molecule. III. Location
            of the P3HR-1 deletion junction and characterization of the NotI
            repeat units that form part of the template for an abundant
            12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript
  JOURNAL   J. Virol. 48 (1), 135-148 (1983)
  MEDLINE   83294686
REFERENCE   6  (bases 159853 to 172281)
  AUTHORS   Bankier,A.T., Deininger,P.L., Satchwell,S.C., Baer,R., Farrell,P.J.
            and Barrell,B.G.
  TITLE     DNA sequence analysis of the EcoRI Dhet fragment of B95-8
            Epstein-Barr virus containing the terminal repeat sequences
  JOURNAL   Mol. Biol. Med. 1 (4), 425-445 (1983)
  MEDLINE   85060428
REFERENCE   7  (bases 1 to 172281)
  AUTHORS   Farrell,P.J., Bankier,A.T., Seguin,C., Deininger,P.L. and
            Barrell,B.G.
  TITLE     Latent and lytic cycle promoters of Epstein-Barr virus
  JOURNAL   EMBO J. 2, 1331-1338 (1983)
REFERENCE   8  (bases 142687 to 159853)
  AUTHORS   Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G.
  TITLE     Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8
            Epstein-Barr virus
  JOURNAL   Mol. Biol. Med. 1 (1), 21-45 (1983)
  MEDLINE   85035713
REFERENCE   9  (bases 112620 to 125316)
  AUTHORS   Seguin,C., Farrell,P.J. and Barrell,B.G.
  TITLE     DNA sequence and transcription of the BamHI fragment B region of
            B95-8 Epstein-Barr virus
  JOURNAL   Mol. Biol. Med. 1 (3), 369-392 (1983)
  MEDLINE   85060424
REFERENCE   10 (bases 45415 to 52824)
  AUTHORS   Jones,M.D., Foster,L., Sheedy,T. and Griffin,B.E.
  TITLE     The EB virus genome in Daudi Burkitt's lymphoma cells has a
            deletion similar to that observed in a non-transforming strain
            (P3HR-1) of the virus
  JOURNAL   EMBO J. 3 (4), 813-821 (1984)
  MEDLINE   84207939
REFERENCE   11 (bases 87650 to 92703)
  AUTHORS   Biggin,M., Farrell,P.J. and Barrell,B.G.
  TITLE     Transcription and DNA sequence of the BamHI L fragment of B95-8
            Epstein-Barr virus
  JOURNAL   EMBO J. 3 (5), 1083-1090 (1984)
  MEDLINE   84236104
REFERENCE   12 (bases 7315 to 9312)
  AUTHORS   Yates,J., Warren,N., Reisman,D. and Sugden,B.
  TITLE     A cis-acting element from the Epstein-Barr viral genome that
            permits stable replication of recombinant plasmids in latently
            infected cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)
  MEDLINE   84222045
REFERENCE   13 (bases 76089 to 79808)
  AUTHORS   Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B.
  TITLE     Homology between two EBV early genes and HSV ribonucleotide
            reductase and 38K genes
  JOURNAL   Nucleic Acids Res. 12 (12), 5087-5099 (1984)
  MEDLINE   84247360
REFERENCE   14 (bases 1 to 172281)
  AUTHORS   Baer,R., Bankier,A.T., Biggin,M.D., Deininger,P.L., Fallell,P.J.,
            Gibson,T.J., Hatfull,G., Hudson,G.S., Satchwell,S.C., Seguin,C.,
            Tuffnell,P.S. and Barrell,B.G.
  TITLE     DNA sequence and expression of the B95-8 Epstein-Barr virus genome
  JOURNAL   Nature 310 (5974), 207-211 (1984)
  MEDLINE   84270667
REFERENCE   15 (bases 1 to 172281)
  AUTHORS   Bodescot,M. and Perricaudet,M.
  TITLE     Clustered alternative splice sites in Epstein-Barr virus RNAs
  JOURNAL   Nucleic Acids Res. 15 (14), 5887 (1987)
  MEDLINE   87289053
REFERENCE   16 (bases 1 to 172281)
  AUTHORS   Laux,G., Perricaudet,M. and Farrell,P.J.
  TITLE     A spliced Epstein-Barr virus gene expressed in immortalized
            lymphocytes is created by circularization of the linear viral
            genome
  JOURNAL   EMBO J. 7 (3), 769-774 (1988)
  MEDLINE   88283646
REFERENCE   17 (bases 1 to 172281)
  AUTHORS   Hatfull,G.F., Barrell,B.G., Quinn,J. and McGeoch,D.
  JOURNAL   Unpublished
REFERENCE   18 (bases 1 to 172281)
  AUTHORS   Farrell,P.J. and Barrell,B.G.
  TITLE     Direct Submission
```

JOURNAL Submitted (05-JUN-1984) to the EMBL/GenBank/DDBJ databases
REFERENCE 19 (bases 1 to 172281)
AUTHORS Farrell,P.J.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1988) Farrell P., Ludwig Institute for Cancer Research, St. Mary's Hospital Medical School, Norfolk Place London W2 1PG
COMMENT CDS

Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In general the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been excluded and on the other hand some small frames have been included which might represent exons or genes because they occur in a logical combination with other features or because of some other experimental data. The reading frames are named according to the Bam H1 fragment in which they start. eg BALF3 is the third leftward frame starting in Bam H1 fragment A. BORF1 is the first rightward frame in Bam H1 fragment O. If there is an obvious TATA sequence followed by an in frame Met codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

SITEs of POLYA signals

This feature lists all occurences of the sequence AATAAA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog ATTAAA is only listed when it is found in a position close to the end of a major reading frame.

SITEs of DONOR and ACCEPT sequences

This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case.

Restriction enzyme SITEs.

Only the positions of the sites Bam HI (BAM) are listed.

RPT

This feature is used to define repetitive sequences.

SITE DEL

This feature defines deletions in B95-8 with respect to other strains such as RAJI and also to deletions in other strains such as P3HR1 and DAUDI with respect to B95-8.

SITE HPN

Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurences noted.

ORGRPL

Denotes the region that encompasses an origin of replication (ori P).[13].

NUMBERING

The DNA sequence of B95-8 EBV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCAGTCTTT. To avoid renumbering the entire sequence, position 1 has benn moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC).

FEATURES Location/Qualifiers
source 1. .172281
/organism="Human herpesvirus 4"
/strain="B95-8"
/db_xref="taxon:10376"
mRNA 58. .272
/note="exon 2 terminal protein RNA"
mRNA 360. .458
/note="exon 3 terminal protein RNA"
misc_feature complement(535)
/note="polyA signal: AATAAA"
mRNA 540. .788
/note="exon 4 terminal protein RNA"
mRNA 871. .951
/note="exon 5 terminal protein RNA"
mRNA 1026. .1196
/note="exon 6 terminal protein RNA"
promoter complement(1192)
/note="TATA: TATAAAT"
mRNA 1280. .1495
/note="exon 7 terminal protein RNA"
promoter complement(1383)
/note="TATA: CATAAAA"
mRNA 1574. .1682
/note="exon 8 terminal protein RNA"
promoter 1676
/note="TATA: TATAAAG"
promoter 1691
/note="TATA: TATTAAA BN-R1 late promoter before BNRF1, gives 4.1kb late RNA. Probably encodes non glycosylated 140kd protein in membrane antigen. Also two latent RNAs spliced underneath this RNA, lengths 1.8 and 2.0kb (Hudson et al, 1985). The longer one encodes terminal protein."
CDS 1736. .5692
/note="BNRF1 reading frame, 5 NXT/S"
/codon_start=1
/db_xref="PID:g59075"
/db_xref="SWISS-PROT:P03179"
/translation="MEERGRETQMPVARYGGPFIMVRLFGQDGEANIQEERLYELLSD
PRSALGLDPGPLIAENLLLVALRGTNNDPRPQRQERARELALVGILLGNGEQGEHLGT
ESALEASGNNYVYAYGPDWMARPSTWSAEIQQFLRLLGATYVLRVEMGRQFGFEVHRS
RPSFRQFQAINHLVLFDNALRKYDSGQVAAGFQRALLVAGPETADTRPDLRKLNEWVF
GGRAAGGRQLADELKIVSALRDTYSGHLVLQPTETLDTWKVLSRDTRTAHSLEHGFIH
AAGTIQANCPQLFMRRQHPGLFPFVNAIASSLGWYYQTATGPGADARAAARRQQAFQT
RAAAECHAKSGVPVVAGFYRTINATLKGGEGLQPTMFNGELGAIKHQALDTVRYDYGH
YLIMLGPFQPWSGLTAPPCPYAESSWAQAAVQTALELFSALYPAPCISGYARPPGPSA
VIEHLGSLVPKGGLLLFLSHLPDDVKDGLGEMGPARATGPGMQQFVSSYFLNPACSNV
FITVRQRGEKINGRTVLQALGRACDMAGCQHYVLGSTVPLGGLNFVNDLASPVSTAEM
MDDFSPFFTVEFPPIQEEGASSPVPLDVDESMDISPSYELPWLSLESCLTSILSHPTV
GSKEHLVRHTDRVSGGRVAQQPGVGPLDLPLADYAFVAHSQVWTRPGGAPPLPYRTWD
RMTEKLLVSAKPGGENVKVSGTVITLGEQGYKVSLDLREGTRLAMAEALLNAACAPIL
DPEDVLLTLHLHLDPRRADNSAVMEAMTAASDYARGLGVKLTFGSASCPETGSSASNF
MTVVASVSAPGEFSGPLITPVLQKTGSLLIAVRCGDGKIQGGSLFEQLFSDVATTPRA
PEALSLKNLFRAVQQLVKSGIVLSGHDISDGGLVTCLVEMALAGQRGVTITMPVASDY
LPEMFAEHPGLVFEVEERSVGEVLQTLRSMNMYPAVLGRVGEQGPDQMFEVQHGPETV
LRQSLRLLLGTWSSFASEQYECLRPDRINRSMHVSDYGYNEALAVSPLTGKNLSPRRL
VTEPDPRCQVAVLCAPGTRGHESLLAAFTNAGCLCRRVFFREVRDNTFLDKYVGLAIG
GVHGARDSALAGRATVALINRFPALRDAILKFLNRPDTFSVALGELGVQVLAGLGAVG
STDNPPAPGVEVNVQRSPLILAPNASGMFESRWLNISIPATTSSVMLRGLRGCVLPCW
VQGSCLGLQFTNLGMPYVLQNAHQIACHFHSNGTDAWRFAMNYPRNPTEQGNIAGLCS
RDGRHLALLCDPSLCTDFWQWEHIPPAFGHPTGCSPWTLMFQAAHLWSLRHGRPSE"

```
misc_feature    complement(1795)
                /note="polyA signal: AATAAA"
misc_feature    3955
                /note="BAM: Bam H1 Nhet/h"
misc_feature    3994
                /note="BAM: Bam H1 h/C"
mRNA            5408. .5856
                /note="exon 9 terminal protein RNA"
misc_feature    5841
                /note="polyA signal: AATAAA, end of 4.1kb late RNA and TP
                latent RNA."
misc_feature    5863
                /note="alternative end to TP cDNAs"
promoter        6097
                /note="TATA: TATAAGA"
misc_feature    6629. .6795
                /note="Pol III RNA EBER 1"
promoter        complement(6823)
                /note="TATA: CATAAAT"
misc_feature    6956. .7128
                /note="Pol III RNA EBER 2"
rep_origin      7315. .9312
                /note="origin of replication, ori P (Yates et al, 1984,
                1985)"
repeat_region   7421. .8042
                /note="21x30bp repeats, binding sites for EBNA-1 (site I,
                Rawlins et al, 1985). Tandem repeat part of oriP (Reisman
                et al, 1985). Also functions as a cell type specific
                enhancer (Reisman et al, 1985; Lupton and Levine, 1985)"
promoter        7738
                /note="TATA: TATAAAT"
promoter        7888
                /note="TATA: TATAAAT"
promoter        8573
                /note="TATA: CATAAAT"
misc_feature    complement(8680)
                /note="polyA signal: AATAAA"
misc_feature    complement(8755)
                /note="polyA signal: AATAAA"
misc_feature    8962
                /note="polyA signal: AATAAA"
misc_feature    9021. .9133
                /note="HPN: dyad symmetry, site II for EBNA-1 binding
                (Rawlins et al, 1985). Dyad symmetry part of oriP (Reisman
                et al, 1985)"
promoter        complement(9398)
                /note="TATA: TATAAAT"
promoter        9631
                /note="TATA: TATAAAT BC-R1 late promoter before BCRF1"
CDS             9675. .10187
                /note="BCRF1 reading frame"
                /codon_start=1
                /db_xref="PID:g59076"
                /db_xref="SWISS-PROT:P03180"
                /translation="MERRLVVTLQCLVLLYLAPECGGTDQCDNFPQMLRDLRDAFSRV
                KTFFQTKDEVDNLLLKESLLEDFKGYLGCQALSEMIQFYLEEVMPQAENQDPEAKDHV
                NSLGENLKTLRLRLRRCHRFLPCENKSKAVEQIKNAFNKLQEKGIYKAMSEFDIFINY
                IEAYMTIKAR"
promoter        10076
                /note="TATA: GATAAAA"
misc_feature    complement(10148)
                /note="polyA signal: AATAAA"
misc_feature    10173
                /note="polyA signal: ATTAAA"
misc_feature    10257
                /note="polyA signal: AATAAA, end of 0.8kb late RNA from
                BCR1 and end of 1.6 kb late RNA, start unknown"
misc_feature    complement(10277)
                /note="polyA signal: AATAAA"
promoter        complement(10975)
                /note="TATA: CATAAAT"
promoter        11305
                /note="TATA: TACAAAA; BCR2 promoter for highly spliced
                EBNA latent RNAs."
mRNA            11336. .11480
                /note="exon C1 of Bodescot et al (1986) RNAs"
promoter        11524
                /note="TATA: TATAATT"
misc_feature    complement(11587)
                /note="polyA signal: AATAAA"
promoter        complement(11606)
                /note="TATA: CATAAAT"
mRNA            11626. .11657
                /note="exon C2 of Bodescot et al (1986) RNAs"
promoter        11796
                /note="TATA: TATAAGT"
promoter        complement(11799)
                /note="TATA: TATAAAA"
repeat_region   12001. .15072
                /note="3072 repeat 1"
CDS             12541. .13689
                /note="BCRF2 3072 repeat, reading frame 1; Protein
                sequence is in conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25077"
                /db_xref="PID:g1334836"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    13215
                /note="BAM: BamH1 C/W"
promoter        14352
                /note="*TATA: TATAAAG BWR1 one of the promoters for highly
                spliced EBNA and LP RNAs (Sample et al, 1986; Speck et al,
                1986)"
mRNA            14384. .14410
                /note="*exon W0 of EBNA/LP RNAs"
mRNA            14554. .14619
                /note="*exon W1 (also W66) part of leader protein (LP)
                gene. LP is also called EBNA-5 (Dillner et al, 1986) and
                EBNA-4 (Rowe et al, 1987)."
mRNA            14559. .14619
                /note="*exon W1' (also W61) of EBNA/LP RNAs forms
                initiator met when fused to exon W0 or exon C2."
mRNA            14701. .14832
                /note="*exon W2 (also W132) part of LP gene"
repeat_region   15073. .18144
                /note="3072 repeat 2"
CDS             15613. .16761
                /note="BWRF1 reading frame 2; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25078"
                /db_xref="PID:g1334837"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    16287
                /note="BAM: BamH1 W/W"
promoter        17424
                /note="TATA: TATAAAG"
mRNA            17626. .17691
                /note="Exon W1"
mRNA            17773. .17904
                /note="Exon W2"
repeat_region   18145. .21216
```

```
                /note="3072 repeat 3"
CDS             18685. .19833
                /note="BWRF1 reading frame 3; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25079"
                /db_xref="PID:g1334838"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    19359
                /note="BAM: BamH1 W/W"
promoter        20496
                /note="TATA: TATAAAG"
mRNA            20698. .20763
                /note="Exon W1"
mRNA            20845. .20976
                /note="Exon W2"
repeat_region   21217. .24288
                /note="3072 repeat 4"
CDS             21757. .22905
                /note="BWRF1 reading frame 4; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25080"
                /db_xref="PID:g1334839"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    22431
                /note="BAM: BamH1 W/W"
promoter        23568
                /note="TATA: TATAAAG"
mRNA            23771. .23835
                /note="Exon W1"
mRNA            23917. .24048
                /note="Exon W2"
repeat_region   24289. .27360
                /note="3072 repeat 5"
CDS             24829. .25977
                /note="BWRF1 reading frame 5; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25081"
                /db_xref="PID:g1334840"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    25503
                /note="BAM: BamH1 W/W"
promoter        26640
                /note="TATA: TATAAAG"
mRNA            26842. .26907
                /note="Exon W1"
mRNA            26989. .27120
                /note="Exon W2"
repeat_region   27361. .30432
                /note="3072 repeat 6"
CDS             27901. .29049
                /note="BWRF1 reading frame 6; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25066"
                /db_xref="PID:g1334841"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    28575
                /note="BAM: BamH1 W/W"
promoter        29712
                /note="TATA: TATAAAG"
mRNA            29914. .29979
                /note="Exon W1"
mRNA            30061. .30192
                /note="Exon W2"
repeat_region   30433. .33504
                /note="3072 repeat 7"
CDS             30973. .32121
                /note="BWRF1 reading frame 7; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25067"
                /db_xref="PID:g1334842"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    31647
                /note="BAM: BamH1 W/W"
promoter        32784
                /note="TATA: TATAAAG"
mRNA            32986. .33051
                /note="Exon W1"
mRNA            33133. .33264
                /note="Exon W2"
repeat_region   33505. .36576
                /note="3072 repeat 8"
CDS             34045. .35193
                /note="BWRF1 reading frame 8; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25068"
                /db_xref="PID:g1334843"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    34719
                /note="BAM: BamH1 W/W"
promoter        35856
                /note="TATA: TATAAAG"
mRNA            36058. .36123
                /note="Exon W1"
mRNA            36205. .36336
                /note="Exon W2"
repeat_region   36577. .39648
                /note="3072 repeat 9"
CDS             37117. .38265
```

```
                /note="BWRF1 reading frame 9; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25069"
                /db_xref="PID:g1334844"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    37791
                /note="BAM: BamH1 W/W"
promoter        38928
                /note="TATA: TATAAAG"
mRNA            39130. .39195
                /note="Exon W1"
mRNA            39277. .39408
                /note="Exon W2"
repeat_region   39649. .42720
                /note="3072 repeat 10"
CDS             40189. .41337
                /note="BWRF1 reading frame 10; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25070"
                /db_xref="PID:g1334845"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    40863
                /note="BAM: BamH1 W/W"
promoter        42000
                /note="TATA: TATAAAG"
mRNA            42202. .42267
                /note="Exon W1"
mRNA            42349. .42480
                /note="Exon W2"
repeat_region   42721. .45792
                /note="3072 repeat 11"
CDS             43261. .44409
                /note="BWRF1 reading frame 11; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25071"
                /db_xref="PID:g1334846"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    43935
                /note="BAM: BamH1 W/W"
promoter        45072
                /note="TATA: TATAAAG"
mRNA            45274. .45339
                /note="Exon W1"
misc_feature    45415. .52824
                /note="DEL: DAUDI deletion (Jones et al, 1984)"
mRNA            45421. .45552
                /note="Exon W2"
misc_feature    45644. .52450
                /note="DEL: P3HR1 deletion (Jeang and Hayward, 1983)"
repeat_region   45793. .47643
                /note="3072 repeat 12"
CDS             46333. .47481
                /note="BWRF1 reading frame 12; Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /db_xref="PID:e25072"
                /db_xref="PID:g1334847"
                /db_xref="SPTREMBL:Q04397"
                /translation="VWEAEGRPRPGEVEGDRPGLCWQSPGDPLRPSGPGRSPSAPQTD
                PRVSRQGPASSGAAGSPPQAPQTRVSASRADRPRAWRLLGASRRGWFCPSLCPSEEPG
                TSGTPEPLGPASRRPPGLRSPLSPVKPKECLRGATLGAQAPESRGQGHLRVPPRVPGQ
                PEGPRQPGRPQRPVPRPFPGLQSPGCPPEGTLGVPSPPLQARASPSRRGASLGPQVQP
                HRDPSGPDPPTGPSLCPPAPLQPSLHPRPQLLASPGPPGQPEGPRQPGRVAFPLPWPL
                LPASHPSPLSLPPHRVHQAGRRDPGGPVSVPPAAAQSLPPGKGASFSPPSLRPSLLCT
                VCKVQPPTPVHGSRAQPRPPLPTVDRPSVHPGHPRPPVSTPVPSRGDFM"
misc_feature    47007
                /note="BAM: BamH1 W/Y"
mRNA            47761. .47793
                /note="Exon Y1 Bodescot et al, 1984"
promoter        47831
                /note="TATA: TATAAGT"
mRNA            47878. .47999
                /note="Exon Y2 Bodescot et al, 1984 and EBNA-1 (Speck and
                Strominger,1985), last common exon"
misc_feature    complement(48023)
                /note="polyA signal: AATAAA"
mRNA            48386. .48444
                /note="exon Bodescot et al, 1984"
CDS             48386. .50032
                /note="Coding exon for EBNA-2 (Sample et al,1986)"
                /codon_start=1
                /db_xref="PID:e25073"
CDS             48504. .49967
                /note="BYRF1, encodes EBNA-2 (Dambaugh et al, 1984;
                Dillner et al, 1984)"
                /codon_start=1
                /db_xref="PID:e25074"
                /db_xref="PID:g1632787"
                /db_xref="SWISS-PROT:P12978"
                /translation="MPTFYLALHGGQTYHLIVDTDSLGNPSLSVIPSNPYQEQLSDTP
                LIPLTIFVGENTGVPPPLPPPPPPPPPPPPPPPPPPPPPPPSPPPPPPPPPPQR
                RDAWTQEPSPLDRDPLGYDVGHGPLASAMRMLWMANYIVRQSRGDRGLILPQGPQTAP
                QARLVQPHVPPLRPTAPTILSPLSQPRLTPPQPLMMPPRPTPPTPLPPATLTVPPRPT
                RPTTLPPTPLLTVLQRPTELQPTPSPPRMHLPVLHVPDQSMHPLTHQSTPNDPDSPEP
                RSPTVFYNIPPMPLPPSQLPPPAAPAQPPPGVINDQQLHHLPSGPPWWPPICDPPQPS
                KTQGQSRGQSRGRGRGRGRGRGKGKSRDKQRKPGGPWRPEPNTSSPSMPELSPVLGLH
                QGQGAGDSPTPGPSNAAPVCRNSHTATPNVSPIHEPESHNSPEAPILFPDDWYPPSID
                PADLDESWDYIFETTESPSSDEDYVEGPSKRPRPSIQ"
repeat_region   48678. .48800
                /note="14 x CCCCCACCA repeats"
misc_feature    48848
                /note="BAM: BamH1 Y/H"
promoter        49350
                /note="TATA: TATAACA"
promoter        complement(49353)
                /note="TATA: TATAAAA"
repeat_region   49525. .49578
                /note="9 x GGGGCA repeats"
mRNA            49852. .50032
                /note="exon (Bodescot et al 1984)"
misc_feature    50003
                /note="polyA signal: AATAAA, end of Bodescot T1 RNA and
                EBNA-2 RNA (3.0kb latent RNA in IB4 cells)"
promoter        complement(50156)
                /note="TATA: TATAAGT"
misc_feature    complement(50317)
                /note="polyA signal: AATAAA, end of 2.5kb early RNA from
                52817"
misc_feature    complement(50578. .52557)
                /note="BHLF1 early reading frame"
repeat_region   50578. .52115
                /note="12 x 125bp repeats"
```

misc_feature 52654. .53697
/note="region homologous to Eco R1 C of Raji"
promoter complement(52817)
/note="TATA: GATAAAA promoter for 2.5kb early RNA containing BHLF1 (Jeang and Hayward, 1983; Freese et al,1983)"
promoter 53759
/note="TATA: TATTAAC likely promoter for class III and IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
misc_feature 53895
/note="DONOR: CGGGTAACT donor for splice to 54335 in class IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
misc_feature 54335
/note="ACCEPT: TTTTCTAG acceptor from 48444 in class I, 47999 in class II, and 53895 in class IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
misc_feature 54376. .54948
/note="BHRF1 reading frame, limited homolgy to bcl-2 gene. Early gene in B95-8 cells and part of restricted EA complex."
promoter 54591
/note="TATA: TATAACA"
promoter complement(54594)
/note="TATA: TATAAAT"
misc_feature 54853
/note="BAM: BamH1 H/F"
misc_feature complement(54929)
/note="polyA signal: AATAAA"
promoter complement(54977)
/note="TATA: TATAAAG"
misc_feature 55518
/note="polyA signal: AATAAA, 3' end of 2.5kb, 1.9kb, 1.7kb and 0.6kb early RNAs"
misc_feature complement(55982. .56935)
/note="BFLF2 reading frame, 4 NXT/S, homologous to RF 27 in VZV and HFRF2 in CMV"
misc_feature complement(55990)
/note="polyA signal: AATAAA, 3' end of 2.3kb and 1.1kb early RNAs from 58568 and 57081"
promoter complement(56132)
/note="TATA: TATAAAG"
CDS complement(56948. .58525)
/note="BFLF1 reading frame, 2 NXT/S homologous to RF 26 in VZV and HFRF1 in CMV"
/codon_start=1
/db_xref="PID:g1334849"
/db_xref="SWISS-PROT:P03184"
/translation="MAHKVTSANEPNPLTGKRLSSCPLTRSGVTEVAQIAGRTPKMED
FVPWTVDNLKSQFEAVGLLMAHSYLPANAEEGIAYPPLVHTYESLSPASTCRVCDLLD
TLVNHSDAPVAFFEDYALLCYYCLNAPRAWISSLITGMDFLHILIKYFPMAGGLDSLF
MPSRILAIDIQLHFYICRCFLPVSSSDMIRNANLGYYKLEFLKSILTGQSPANFCFKS
MWPRTTPTFFLTLPGPRTCKDSQDVPGDVGRGLYTALCCHLPTRNRVQHPFLRAEKGGL
SPEITTKADYCGLLLGTWQGTDLLGGPGHHAIGLNAEYSGDELAELALAITRPEAGDH
SQGPCLLAPMFGLRHKNASRTICPLCESLGAHPDAKDTLDRFKSLILDSFGNNIKILD
RIVFLIKTQNTLLDVPCPRLRAWLQMCTPQDFHKHLFCDPLCAINHSITNPSVLFGQI
YPPSFQAFKAALAAGQNLEQGVCDSLITLVYIFKSTQVARVGKTILVDVTKELDVVLR
IHGLDLVQSYQTSQVYV"
promoter complement(57081)
/note="TATA: TATTTAA before BFLF2; BFL2 promoter gives 1.1kb early RNA"
promoter complement(58088)
/note="TATA: GATAAAA"
promoter complement(58568)
/note="TATA: TATTAAA before BFLF1, BFL1 promoter gives 2.3kb early RNA"
promoter 58832
/note="TATA: TATAAAA before BFRF1"
CDS 58891. .59901
/note="BFRF1 early reading frame, 1 NXT/S, homologous to HFLF4 in CMV"
/codon_start=1
/db_xref="PID:g1334850"
/db_xref="SWISS-PROT:P03185"
/translation="MASPEERLLDELNNVIVSFLCDSGSLEVERCSGAHVFSRGSSQP
LCTVKLRHGQIYHLEFVYKFLAFKLKNCNYPSSPVFVISNNGLATTLRCFLHEPSGLR
SGQSGPCLGLSTDVDLPKNSIIMLGQDDFIKFKSPLVFPAELDLLKSMVVCRAYITEH
RTTMQFLVFQAANAQKASRVMDMISDMSQQLSRSGQVEDTGARVTGGGGPRPGVTHSG
CLGDSHVRGRGGWDLDNFSEAETEDEASYAPWRDKDSWSESEAAPWKKELVRHPIRRH
RTRETRRMRGSHSRVEHVPPETRETVVGGAWRYSWRATPYLARVLAVTAVALLLMFLR
WT"
CDS 59808. .61583
/note="BFRF2 early reading frame, homologous to HFLF5 in CMV"
/codon_start=1
/db_xref="PID:e25058"
/db_xref="PID:g1632788"
/db_xref="SWISS-PROT:P14347"
/translation="MALFLARHTLSGTGAGCHGRGPAPDVSEVDLTLQALGERGFSRL
LDLGLACLDLSYVEMREFVVWGRPPASEAAVASTPGSLFRSHSSAYWLSEVERPGGLV
RWARSQTSPSSLTLAPHLGPSLLSLSVVTGGGCGAVAFCNAFFLAYFLVVRSVFPAFS
DRIAAWICDRSPFCENTRAVARGYRGLVKRFLAFVFERSSYDPPLLRQNSRPVERCFA
IKNYVPGLDSQSCVTVPSFSRWAQSHASELDPREIRDRVTPATAPSFVADHASALLAS
LQKKASDTPCGNPIQWMWYRLLVNSCLRSAHCLLPIPAVSEGGRKTGGGVGEELVGAG
GPCLSRDVFVAIVSRNVLSCLLNVPAAGPRAYKCFRSHASRPVSGPDYPPLAVFCMDC
GYCLNFGKQTGVGGRLNSFRPTLQFYPRDQKEKHVLTCHASGRVYCSNCGSAAVGCQR
LAEPPSARSGWRPRIRAVLPHNAAYELDRGSRLLDAIIPCLGPDRTCMRPVVLRGVTV
RQLLYLTLRTEARAVCSICQQRQAPEDARDEPHLFSSCLEVELPPGERCAGCRLYQTR
YGTPAAQAHPPGEAGGGFSRQSPAS"
promoter complement(61062)
/note="TATA: GATAAAA"
promoter 61344
/note="TATA: TATTTAA before BFRF3"
CDS 61507. .62037
/note="BFRF3 early reading frame"
/codon_start=1
/db_xref="PID:e25059"
/db_xref="PID:g1632789"
/db_xref="SWISS-PROT:P14348"
/translation="MARRLPKPTLQGRLEADFPDSPLLPKFQELNQNNLPNDVFREAQ
RSYLVFLTSQFCYEEYVQRTFGVPRRQRAIDKRQRASVAGAGAHAHLGGSSATPVQQA
QAAASAGTGALASSAPSTAVAQSATPSVSSSISSLRAATSGATAAASAAAAVDTGSGG
GGQPHDTAPRGARKKQ"
misc_feature complement(62068)
/note="polyA signal: AATAAA"
misc_feature 62069
/note="polyA signal: AATAAA, 3' end of 10, 6.5, 3.7, 3.4, 3.1, 2.5 and 0.8kb early RNAs"
CDS complement(62078. .71527)
/note="BPLF1 reading frame, 1 NXT/S, analogous to VZV RF22"
/codon_start=1
/db_xref="PID:g1334853"
/db_xref="SWISS-PROT:P03186"
/translation="MSNGDWGQSQRTRGTGPVRGIRTMDVNAPGGGSGGSALRILGTA
SCNQAHCKFGRFAGIQCVSNCVLYLVKSFLAGRPLTSRPELDEVLDEGARLDALMRQS
GILKGHEMAQLTDVPSSVVLRGGGRVHIYRSAEIFGLVLFPAQIANSAVVQSLAEVLH
GSYNGVAQFILYICDIYAGAIIIETDGSFYLFDPHCQKDAAPGTPAHVRVSTYAHDIL
QYVGAPGAQYTCVHLYFLPEAFETEDPRIFMLEHYGVYDFYEANGSGFDLVGPELVSS
DGEAAGTPGADSSPPVMLPFERRIIPYNLRPLPSRSFTSDSFPAARYSPAKTNSPPSS
PASAAPASAAPASAAPASAAPASAAPASAAPASAAPASAAPASSPPLFIPIPGLGHTP
GVPAPSTPPRASSGAAPQTPKRKKGLGKDSPHKKPTSGRRLPLSSTTDTEDDQLPRTH
VPPHRPPSAARLPPPVIPIPHQSPPASPTPHPAPVSTIAPSVTPSPRLPLQIPIPLPQ
AAPSNPKIPLTTPSPSPTAAAAPTTTTLSPPPTQQQPPQSAAPAPSPLLPQQQPTPSA
APAPSPLLPQQQPPPSAARAPSPLPPQQQPLPSATPAPPPAQQLPPSATTLEPEKNHP
PAADRAGTEISPSPPFGQQPSFGDDASGGSGLVRYLSDLEEPFLSMSDSEEAESDLAS
DIPTTEDEDMFEDEVFSNSLESGSSAPTSPITLDTARSQYYQTTFDIETPEMDFVPLE
SNIARIAGHTYQEQAIVYDPASNREVPEADALSMIDYLLVTVVLEQGLIRSRDRSSVL
NLLEFLKDWSGHLQVPTLDLEQLLTSELNIQNLANMLSENKGRAGEFHKHLAAKLEAC
LPSLATKDAVRVDAGAKMLAEIPQLAESDDGKFDLEAARRRLTDLLSGGDQEAGEGGG
EPEDNSIYRGPHVDVPLVLDDESWKRLLSLAEAARTAVARQQAGVDEEDVRFLALLTA
IEYGAPPAASVPPFVHNVAVRSKNAALHVRRCTADIRDKVASAASDYLSYLEDPSLPT
VMDFDDLLTHLRHTCQIIASLPLLNIRYTSIEWDYRELLYLGTALSDMSGIPWPLERV
EEDDPSIAPLPEFETVAKKQKELETTRENEKRLRTILDDIEAMLGLAGVASAPGAPIS
PASPSATPANHDNPEATPPLADTAALTIPVIEKYIANAGSIVGAAKNPTYIRLRDTIQ
QIVRSKKYLMNILKSITFYTIDNYIASFEESIDHLYRDLPVLDPEVQDGIDRILDPMV
SEALHTFEMGNRLTLEPARLVALQNFATHSTLKETAAAVNLLPGLLAVYDATITGQAP

EDALRLLSGLQNQLSQTLIPGKLKKRFLSYLQKLKNNNNDQLRQKEVQAWRLEAEGFK
PATEEQLEAFLDTAPNKELKRQYEKKLRQLMETGRKEKEKLREQEDKERQERRAREAN
EAWARIRKALGARPEPAPTSPDDWNTLLASLLPDNTDSAAAAAAAVARNTDILDSLTQ
ILAAMLLGITRVRRERLRSLLVDDGGAAERMEAAEPGWFTDIETGPLARLDAWPATPA
ATAKEGGGGRGAEEAAGALFRARTAADAIRSALAQTRQALQSPDMKSAVVNTDLEAPY
AEYERGLAGLLEKRRAAEAALTAIVSEYVDRTLPEATNDPGQANLPPPPTIPQATAPP
RLASDSALWPKKPQLLTRRERDDLLQATGDFFSELLTEAEAAEVRALEEQVRESQTLM
AKAHEMAASTRRGFHTALEAVLSRSRDEAPDDELRSLLPSPPKAPVQAPLEAALARAA
AGNGSWPYRKSLAAAKWIRGICEAVRGLSEGALALAGGAGAWLNLAAAADGEIHELTR
LLEVEGMAQNSMDGMEELRLALATLDPKRVAGGKETVADWKRRLSRLEAIIQEAQEES
QLQGTLQDLVTQARGHTDPRQLKIVVEAARGLALGASAGSQYALLKDKLLRYASAKQS
FLAFYETAQPTVFVKHPLTNNLPLLITISAPPTGWGNGAPTRRAQFLAAAGPAKYAGT
LWLETESPCDPLNPAYVSADTQEPLNYIPVYHNFLEYVMPTVLENPEAFSLTPAGRPQ
AIGPPQDDQERRRRTLASVASARLSAAAADSYWDTWPDVESNAGELLREYVSAPKALM
EDLADNPIVAMTLLAHASLIASRNHPPYPAPATDREVILLEQREMMALLVGTHPAYAA
AFLGAPSFYAGLGLVSALARDGGLGDLLSDSVLTYRLVRSPASGRGGMPSTTRGSNDG
EDARRLTRHRIAGPPTGFIFFQDAWEEMDTRAALWPHPEFLGLVHNQSTARARACMLL
LARRCFAPEALQQLWHSLRPLEGPVAFQDYLRDFVKQAYTRGEELPRAEGLEVPRETP
SSYGTVTGRALRNLMPYGTPITGPKRGSGDTIPVSVFEAAVAAAFLGRPLTLFVSSQY
LFNLKTLGQVRVVAPLLYCDGHSEPFRSLVETISLNFLQDLDGYSESFEPEMSIFARQ
AVWLRELLTEARAAKPKEARPPTVAILANRKNIIWKCFTYRHNLPDVQFYFNAAGASR
WPTDVLNPSFYEHEDPPLPVGYQLPPNPRNVQELFSGFPPRVGHGLVSGDGFQSADNT
PASSDRLQQLGGGETDQGEKGSTTAESEASGPPSPQSPLLEKVAPGRPRDWLSPTSSP
RDVTVTPGLAAPITLPGPRLMARPYFGAETRASESPDRSPGSSPRPWPKDSLELLPQP
APQQPPSSPWASEQGPIVYTLSPHSTPSTASGSQKKHTIQIPGLVPSQKPSYPPSAPY
KPGQSTGGIAPTPSAASLTTFGLQPQDTQASSQDPPYGHSIMQREKKQQGGREEAAEI
RPSATRLPTAVGLRPRAPVVAAGAAASATPAFDPGEAPSGFPIPQAPALGSGLAAPAH
TPVGALAPRPQKTQAQRPQDAAALPTPTIKAVGARPVPKATGALAAGARPRGQPTAAP
PSAASPPRVSLPVRSRQQQSPAIPLPPMHSGSEPGARPEVRLSQYRHAGPQTYTVRKE
APPSAASQLPKMPKCKDSMYYPPSGSARYPAPFQALSFSQSVASPAPSSDQTTLLWNT
PSVVTQFLSIEDIIREVVTGGSTSGDLVVPSGSPSSLSTAAPEQDLRYSLTLSQASRV
LSRFVSQLRRKLERSTHRLIADLERLKFLYL"

misc_feature 62249
/note="BAM: BamH1 F/Q"

misc_feature 62430. .62477
/note="Site III for EBNA-1 binding (Rawlins et al, 1985)"

misc_feature 66121
/note="BAM: BamH1 Q/U"

mRNA 67477. .67649
/note="Exon in EBNA-1 RNA (Speck and Strominger,1985) and cDNA clone T4 (Bodescot et al, 1986)"

misc_feature 69410
/note="BAM: BamH1 U/P"

repeat_region 69684. .69930
/note="5 x 51bp repeats"

repeat_region 70387. .70521
/note="9 x 15bp repeat"

promoter 70750
/note="TATA: CATAAAA"

CDS complement(71520. .75239)
/note="BOLF1 reading frame, 1 NXT/S analogous to VZV RF 21"
/codon_start=1
/db_xref="PID:g1334855"
/db_xref="SWISS-PROT:P03189"
/translation="MASAMESDSSGGSGGADAQPPLAEVDGGLARVTRQLLLSGDDPA
ARLRALMPLELGIFGLGDLAQPVLVRDFLNTLTLMSGHAYPAAVLRHHAYYLLRAASF
SRRSFGLGHLEAALDVLASSLPPTTASPATDDPLDGSRLIAETRALAAAYRRIIEEGS
GEVLAVSGPTATFAFVEELVADTYLARWDAFPREGLSFYAFNAAKTTLGRWLVTVYAE
TNRYPWAAAGQGQPTAADIKAMAVELVEHSGGGAGGGEGEESGGGLFHRPESLSSVVA
SLPLARRRAVEILGVYAEASGGQTPPVAAVPVLAFDAARLRLLEPSGALFYDYVYEAL
LWDQTYGVPDSVIEAFLAGMAAEMEALAARVQEAAGSRASFSPAAIEQVATVLLSAGL
NETVAGDYAMMLASVPRVSRSRWRWLEATAALLESLSGFALHFFRLLPTASPTSRFAR
VARAAYLRAEAEAVDRRARRTSGPSTPAAAPAATAVGVGAAADPWDAVTPLRIFIVPP
PAAEYEQVAGDLSSELLRSLLWVRYSRLWQAPAPAPALPCKPPLLPGEQGRRQWTAAV
AAAPRTDVEAYCRSLRAGQTARADPAYVHSPFFPAAFIEFQIWPALRRVLSNELPKTR
SLAALRWLVSFGSDLALPSPELTRARRPLELIYATVWEIYDGAPPMPGESPQAVGLRP
LNLEGEGKAGDAGAEGAEDEEGGGPWGLSSHDAVLRIMDAVREVSGIISETISASERA
AEAPPLAWPTSLFSLLFTLRYSTTAESLGLATRRFLVSGETLSEDISRLTGAAWRLCS
RPLLYDAETGRVQIPLATEEEEEAVVAVKEKSVSSSPRHYSTDLQTLKSVVEGIQDVC
RDAAARWALATADTATLRRRLLVPALRESRGIADHPLWAHTSEPLRPDLEELNERVEH
ALELGYSLTGALRRSVAYRFRDYTFARLFQPPAIDAERAEAIVRRDARPPPVFIPAPR
RLPQGGADTPPPLSMDDILYLGKSICKALVDVLDHHPAAPETTPIKTYTPAMDLNPEQ
ITVTPRSPSVLAAFARTARVQTHHLVPALTDDSPSPVGQTPPPFRILPAKKLAAILLG
NGRNASKRRASRDLSPPPHGRWRAVLDSSPFSFSSSDFSDQDEGEGGEADLRGVPGGG
GEGAYEEDRERPSDIDTAARAQKVETSCPRRRSPRTTPSPSRRASGGGGPDRGEAEAH
TYPPYLSAAAAASRVRPRTRRGATRRPPRPTAEDE"

promoter complement(72192)
/note="TATA: TATTAAA before BPLF1"

misc_feature 73468
/note="BAM: BamH1 P/O"

promoter 75017
/note="TATA: TATTTAA BO-R1 late promoter before BORF1, gives 3.9kb late RNA"

CDS 75238. .76332
/note="BORF1 late reading frame, 2 NXT/S homologous to VZV RF20"
/codon_start=1
/db_xref="PID:g1334854"
/db_xref="SWISS-PROT:P03187"
/translation="MKVQGSVDRRRLQRRIAGLLPPPARRLNISRGSEFTRDVRGLVE
EHAQASSLSAAAVWRAGLLAPGEVAVAGGGSGGGSFSWSGWRPPVFGDFLIHASSFNN
AEATGTPLFQFKQSDPFSGVDAVFTPLSLFILMNHGRGVAARVEAGGGLTRMANLLYD
SPATLADLVPDFGRLVADRRFHNFITPVGPLVENIKSTYLNKITTVVHGPVVSKAIPR
STVKVTVPQEAFVDLDAWLSGGAGGGGGVCFVGGLGLQPCPADARLYVALTYEEAGPR
FTFFQSSRGHCQIMNILRIYYSPSIMHRYAVVQPLHIEELTFGAVACLGTFSATDGWR
RSAFNYRGSSLPVVEIDSFYSNVSDWEVIL"

promoter complement(75322)
/note="TATA: TATTTAG before BOLF1"

promoter 75819
/note="TATA: TATAAAG"

misc_feature 75838
/note="polyA signal: AATAAA"

misc_feature complement(76126)
/note="polyA signal: AATAAA"

promoter 76169
/note="TATA: TACATAT BO-R2 early promoter before BORF2, gives 2.8kb RNA"

misc_feature complement(76300)
/note="polyA signal: AATAAA"

CDS 76407. .78887
/note="BORF2 early reading frame, 2 NXT/S. Homology HSV 140K ribonucleotide reductase (Gibson et al, 1984) and RF 19 VZV"
/codon_start=1
/db_xref="PID:g1334856"
/db_xref="SWISS-PROT:P03190"
/translation="MATTSHVEHELLSKLIDELKVKANSDPEADVLAGRLLHRLKAES
VTHTVAEYLEVFSDKFYDEEFFQMHRDELETRVSAFAQSPAYERIVSSGYLSALRYYD
TYLYVGRSGKQESVQHFYMRLAGFCASTTCLYAGLRAALQRARPEIESDMEVFDYYFE
HLTSQTVCCSTPFMRFAGVENSTLASCILTTPDLSSEWDVTQALYRHLGRYLFQRAGV
GVGVTGAGQDGKHISLLMRMINSHVEYHNYGCKRPVSVAAYMEPWHSQIFKFLETKLP
ENHERCPGIFTGLFVPELFFKLFRDTPWSDWYLFDPKDAGDLERLYGEEFEREYYRLV
TAGKFCGRVSIKSLMFSIVNCAVKAGSPFILLKEACNAHFWRDLQGEAMNAANLCAEV
LQPSRKSVATCNLANICLPRCLVNAPLAVRAQRADTQGDELLLALPRLSVTLPGEGAV
GDGFSLARLRDATQCATFVVACSILQGSPTYDSRDMASMGLGVQGLADVFADLGWQYT
DPPSRSLNKEIFEHMYFTALCTSSLIGLHTRKIFPGFKQSKYAGGWFHWHDWAGTDLS
IPREIWSRLSERIVRDGLFNSQFIALMPTSGCAQVTGCSDAFYPFYANASTKVTNKEE
ALRPNRSFWRHVRLDDREALNLVGGRVSCLPEALRQRYLRFQTAFDYNQEDLIQMSRD
RAPFVDQSQSHSLFLREEDAARASTLANLLVRSYELGLKTIMYYCRIEKAADLGVMEC
KASAALSVPREEQNERSPAEQMPPRPMEPAQVAGPVDIMSKGPGEGPGGWCVPGGLEV
CYKYRQLFSEDDLLETDGFTERACESCQ"

misc_feature 77835
/note="BAM: Bam H1 O/a"

promoter 78804
/note="TATA: TATAAGT Ba-R1 early promoter before BaRF1, gives 3.5kb RNA"

misc_feature 78883
/note="polyA signal: AATAAA, end of 3.9kb late RNA from 75017 and 2.8kb early RNA from 76169"

misc_feature complement(78896)
/note="polyA signal: AATAAA"

CDS 78900. .79808
/note="BaRF1 early reading frame, homologous to HSV 38K ribonucleotide reductase (Gibson et al, 1984) and RF 18 VZV"

```
                /codon_start=1
                /db_xref="PID:g1334857"
                /db_xref="SWISS-PROT:P03175"
                /translation="MSKLLYVRDHEGFACLTVETHRNRWFAAHIVLTKDCGCLKLLNE
                RDLEFYKFLFTFLAMAEKLVNFNIDELVTSFESHDIDHYYTEQKAMENVHGETYANIL
                NMLFDGDRAAMNAYAEAIMADEALQAKISWLRDKVAAAVTLPEKILVFLLIEGIFFIS
                SFYSIALLRVRGLMPGICLANNYISRDELLHTRAASLLYNSMTAKADRPRATWIQELF
                RTAVEVETAFIEARGEGVTLVDVRAIKQFLEATADRILGDIGQAPLYGTPPPKDCPLT
                YMTSIKQTNFFEQESSDYTMLVVDDL"
promoter        complement(79495)
                /note="TATA: TATAACA"
misc_feature    79537
                /note="BAM: Bam H1 a/M"
promoter        79840
                /note="TATA: CATAAAT BM-R1 early promoter before BMRF1,
                gives 2.5kb RNA"
CDS             79899. .81113
                /note="BMRF1 early reading frame. Early antigen protein
                recognised by R3 monoclonal (Pearson et al 1983; Cho et
                al, 1985a)"
                /codon_start=1
                /db_xref="PID:g1334858"
                /db_xref="SWISS-PROT:P03191"
                /translation="METTQTLRFKTKALAVLSKCYDHAQTHLKGGVLQVNLLSVNYGG
                PRLAAVANAGTAGLISFEVSPDAVAEWQNHQSPEEAPAAVSFRNLAYGRTCVLGKELF
                GSAVEQASLQFYKRPQGGSRPEFVKLTMEYDDKVSKSHHTCALMPYMPPASDRLRNEQ
                MIGQVLLMPKTASSLQKWARQQGSGGVKVTLNPDLYVTTYTSGEACLTLDYKPLSVGP
                YEAFTGPVAKAQDVGAVEAHVVCSVAADSLAAALSLCRIPAVSVPILRFYRSGIIAVV
                AGLLTSAGDLPLDLSVILFNHASEEAAASTASEPEDKSPRVQPLGTGLQQRPRHTVSP
                SPSPPPPPRTPTWESPARPETPSPAIPSHSSNTALERPLAVQLARKRTSSEARQKQKH
                PKKVKQAFNPLI"
promoter        80779
                /note="TATA: TATTTAA BM-R2 late promoter before BMRF2"
misc_feature    complement(80782)
                /note="polyA signal: AATAAA"
promoter        80832
                /note="TATA: GATAAAA, possible promoter for 1.4kb late RNA
                encoding BMRF2"
CDS             81118. .82191
                /note="BMRF2 early reading frame"
                /codon_start=1
                /db_xref="PID:g1334859"
                /db_xref="SWISS-PROT:P03192"
                /translation="MFSCKQHLSLGACVFCLGLLASTPFIWCFVFANLLSLEIFSPWQ
                THVYRLGFPTACLMAVLWTLVPAKHAVRAVTPAIMLNIASALIFFSLRVYSTSTWVSA
                PCLFLANLPLLCLWPRLAIEIVYICPAIHQRFFELGLLLACTIFALSVVSRALEVSAV
                FMSPFFIFLALGSGSLAGARRNQIYTSGLERRRSIFCARGDHSVASLKETLHKCPWDL
                LAISALTVLVVCVMIVLHVHAEVFFGLSRYLPLFLCGAMASGGLYLGHSSIIACVMAT
                LCTLTSVVVYFLHETLGPLGKTVLFISIFVYYFSGVAALSAAMRYKLKKFVNGPLVHL
                RVVYMCCFVFTFCEYLLVTFIKS"
promoter        81751
                /note="TATA: CATAAAT"
misc_feature    82180
                /note="polyA signal: ATTAAA, end of 3.5kb early RNA from
                78804, 2.5kb early RNA from 79840 and 1.4kb late RNA"
promoter        complement(82311)
                /note="TATA: CATAAAT"
repeat_region   82319. .82461
                /note="2x71bp repeats"
CDS             complement(82743. .84059)
                /note="BMLF1 early reading frame. Diffuse early antigen
                (Cho et al, 1985b). Also homologous to RF 4 VZV and IE63
                of HSV.(BSLF2 + BMLF1) is also called EB2
                (Chevallier-Greco et al, 1986). General transactivator of
                transcription (Lieberman et al, 1986)."
                /codon_start=1
                /db_xref="PID:e25042"
                /db_xref="PID:g1632790"
                /db_xref="SWISS-PROT:Q04360"
                /translation="MEGSEEHSTDGEISSSEEEDEDPTPAHAIPARPSSVVITPTSAS
                FVIPRKKWDLQDKTVTLHRSPLCRDEDEKEETGNSSYTRGHKRRRGEVHGCTDESYGK
                RRHLPPGARAPRAPRAPRVPRAPRSPRAPRSNRATRGPRSESRGAGRSTRKQARQERS
                QRPLPNKPWFDMSLVKPVSKITFVTLPSPLASLTLEPIQDPFLQSMLAVAAHPEIGAW
                QKVQPRHELRRSYKTLREFFTKSTNKDTWLDARMQAIQNAGLCTLVAMLEETIFWLQE
                ITYHGDLPLAPAEDILLACAMSLSKVILTKLKELAPCFLPNTRDYNFVKQLFYITCAT
                ARQNKVVETLSSSYVKQPLCLLAAYAAVAPAYINANCRRRHDEVEFLGHYIKNYNPGT
                LSSLLTEAVETHTRDCRSASCSRLVRAILSPGTGSLGLFFVPGLNQ"
misc_feature    complement(82747)
                /note="polyA signal: AATAAA"
repeat_region   83640. .83729
                /note="10x9bp repeats"
misc_feature    complement(84122)
                /note="ACCEPT: CTCCCCTCTGCAG acceptor in spliced form of
                BMLF1 RNA"
misc_feature    complement(84227)
                /note="DONOR: CAGGTAAGA donor in spliced form of BMLF1
                RNA"
CDS             complement(84229. .84288)
                /note="BSLF2 early reading frame in 5' exon of spliced RNA
                encoding BMLF1"
                /codon_start=1
                /db_xref="PID:g1334861"
                /db_xref="SPTREMBL:Q66539"
                /translation="MVPSQRLSRTSSISSNEDPA"
misc_feature    84233
                /note="BAM: Bam H1 M/S"
CDS             complement(84257. .86881)
                /note="BSLF1 reading frame, homologous to RF 6 VZV"
                /codon_start=1
                /db_xref="PID:g1334862"
                /db_xref="SWISS-PROT:P03193"
                /translation="MSAPVVIKALVASNTDIAEAILDAILSRPDEGFRLFCLCHNASP
                LHHVAGSLVELQLHLPKKRLTSQSRCGLVLTLHLPAEEAFPFLRGLTPLTADRLSTYL
                DRAGALRSLTPLVELLTLSAKKQPQGDARGRVAWLRPKIVGCLRRIYRVNISARWFIS
                TFGSHEAQFVLVTAAYYFWGIPCTIETLAHLTELFTSESGQSLAAVTSLAELGEVFGS
                SAWAEQTEAFAHFAHEKLRRDSREIRAVARTIDAYRGRLPLASADLVRYVYLAHAQCF
                NEGTFKRYSQLTSMGEIGCLPSGGVVLPSLLDRGFAEHMRTYFTRETYLAEHVRVQQL
                KIRMEPPAPYTWDPDPDDGLMRAWAGLSVDVARELVELARWHADEGPTYPPTLQGFLC
                LAGQATCRGQWNPKEQFLPPTVLRRVQRLPVFLCHFADRHYFVMTAADPFSSHLAEVV
                STPTNCRLPDTCLTRALSYTPVYYSQNSLSEQLFVSRHEYFNPRLPVCNLVLDLDLKI
                KGAPWSLEEIYDLCRTVRREVLRLMRRLGPVSRAHPVYFFKSACPPADPDNMEDVLPF
                CICTGKLGFRVITPLPRGHAIVGTSAVQGFVSVLQKLMGLTACLRRMRHKIKEIGAPL
                FDSGVYHAGRCIRLPHTYKVDRGGGLSRQLRLFVCHPEEEDKHSYVKNALNIQNLLHH
                SLHVGWPAPKTFCYHIADDGRDYLIQRTRETLPPTVENVCAMIEGHLGLDLVAWVSSC
                IWPSLMSTLATAVPEDKFPQFLHVTFEQTGPNLVQVCHARGRNFACLRHTHRASSKNV
                RVFLVLYYTSQAITVTFMSQCFAGRCGANQPTAHFSISVPASRIINRAEASQDSTTSQ
                LARRRDRQDGSFSETLPN"
promoter        complement(84356)
                /note="TATA: CATAAAT before BSLF2 and BMLF1. Two RNAs
                start here; one is spliced and the other is unspliced,
                both traverse BMLF1."
promoter        86882
                /note="TATA: TATTTAA BS-R1 late promoter before BSRF1"
CDS             86924. .87580
                /note="BSRF1 reading frame"
                /codon_start=1
                /db_xref="PID:g1334863"
                /db_xref="SWISS-PROT:P03194"
                /translation="MAFYLPDWSCCGLWLFGRPRNRYSQLPEEPETFECPDRWRAEID
                LGLPPGVQVGDLLRNEQTMGSLRQVYLLAVQANSITDHLKRFDAVRVPESCRGVVEAQ
                VAKLEAVRSVIWNTMISLAVSGIEMDENGLKALLDKQAGDSLALMEMEKVATALKMDE
                TGAWAQEISAVVSSVTAPSASAPFINSAFEPEVPTPVLAPPPVVRQPEHSGPTELALT
                "
misc_feature    complement(87134)
                /note="polyA signal: AATAAA"
misc_feature    87599
                /note="polyA signal: AATAAA"
misc_feature    complement(87613)
                /note="polyA signal: AATAAA, end 1.0kb early RNA from
                BLL3"
CDS             complement(87638. .88474)
                /note="BLLF3 early reading frame (BLLF2 in Baer et al,
                1984). Homologous to RF 8 VZV and dUTPase HSV."
                /codon_start=1
                /db_xref="PID:g1334864"
                /db_xref="SWISS-PROT:P03195"
```

/translation="MEACPHIRYAFQNDKLLLQQASVGRLTLVNKTTILLRPMKTTTV
DLGLYARPPEGHGLMLWGSTSRPVTSHVGIIDPGYTGELRLILQNQRRYNSTLRPSEL
KIHLAAFRYATPQMEEDKGPINHPQYPGDVGLDVSLPKDLALFPHQTVSVTLTVPPPS
IPHHRPTIFGRSGLAMQGILVKPCRWRRGGVDVSLTNFSDQTVFLNKYRRFCQLVYLH
KHHLTSFYSPHSDAGVLGPRSLFRWASCTFEEVPSLAMGDSGLSEALEGRQGRGFGSS
GQ"

misc_feature 87650
/note="BAM: Bam H1 S/L"

promoter 88507
/note="TATA: TATATAT BL-R1 late promoter before BLRF1, gives 1.0kb late RNA"

promoter 88511
/note="TATA: TATAAGA"

promoter complement(88514)
/note="TATA: TATATAT BL-L3 early promoter before BLLF3, gives 1.0kb early RNA"

CDS 88547. .88855
/note="BLRF1 late reading frame"
/codon_start=1
/db_xref="PID:g1334865"
/db_xref="SWISS-PROT:P03196"
/translation="MGKVLRKPFAKAVPLLFLAATWLLTGVLPAGASSPTNAAAASLT
EAQDQFYSYTCNADTFSPSLTSFASIWALLTLVLVIIASAIYLMYVCFNKFVNTLLTD
"

promoter 88863
/note="TATA: TATTTAA BL-R2 late promoter before BLRF2, gives 0.6kb late RNA"

CDS 88925. .89413
/note="BLRF2 late reading frame, 2 NXS/T"
/codon_start=1
/db_xref="PID:g1334866"
/db_xref="SWISS-PROT:P03197"
/translation="MSAPRKVRLPSVKAVDMSMEDMAARLARLESENKALKQQVLRGG
ACASSTSVPSAPVPPPEPLTARQREVMITQATGRLASQAMKKIEDKVRKSVDGVTTRN
EMENILQNLTLRIQVSMLGAKGQPSPGEGTRPRESNDPNATRRARSRSRGREAKKVQI
SD"

misc_feature 89412
/note="polyA signal: AATAAA, end of 1.0kb and 0.6kb late RNAs"

misc_feature complement(89425)
/note="polyA signal: AATAAA, end of 0.7kb early, 2.2kb late and 2.8kb late RNA"

CDS complement(89430. .92153)
/note="BLLF1a, late reading frame, gp350 membrane antigen, 36 NXT /S (Hummel et al, 1984; Biggin et al, 1984; Beisel et al, 1985)"
/codon_start=1
/db_xref="PID:g1334869"
/db_xref="SWISS-PROT:P03200"
/translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTS
STSPVTTPTPNATSPTPAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPT
PNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQAN
ATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSD
NSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGPGNSSTST
KPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTE
PTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRF
SNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"

CDS complement(89430. .92153)
/note="BLLF1b, late reading frame gp220 membrane antigen, spliced form of BLLF1a (Hummel et al, 1984; Biggin et al, 1984; Beisel et al, 1985)"
/codon_start=1
/db_xref="PID:g1334868"
/db_xref="SWISS-PROT:P03200"
/translation="MEAALLVCQYTIQSLIHLTGEDPGFFNVEIPEFPFYPTCNVCTA
DVNVTINFDVGGKKHQLDLDFGQLTPHTKAVYQPRGAFGGSENATNLFLLELLGAGEL
ALTMRSKKLPINVTTGEEQQVSLESVDVYFQDVFGTMWCHHAEMQNPVYLIPETVPYI
KWDNCNSTNITAVVRAQGLDVTLPLSLPTSAQDSNFSVKTEMLGNEIDIECIMEDGEI
SQVLPGDNKFNITCSGYESHVPSGGILTSTSPVATPIPGTGYAYSLRLTPRPVSRFLG
NNSILYVFYSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTNTTDITYVGDNATYSVP
MVTSEDANSPNVTVTAFWAWPNNTETDFKCKWTLTSGTPSGCENISGAFASNRTFDIT
VSGLGTAPKTLIITRTATNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPS
STHVPTNLTAPASTGPTVSTADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTS
STSPVTTPTPNATSPTPAVTTPTPNATSPTPAVTTPTPNATSPTLGKTSPTSAVTTPT
PNATSPTLGKTSPTSAVTTPTPNATSPTLGKTSPTSAVTTPTPNATGPTVGETSPQAN
ATNHTLGGTSPTPVVTSQPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSD
NSTSHMPLLTSAHPTGGENITQVTPASISTHHVSTSSPAPRPGTTSQASGPGNSSTST
KPGEVNVTKGTPPQNATSPQAPSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTE
PTTDYGGDSTTPRPRYNATTYLPPSTSSKLRPRWTFTSPPVTTAQATVPVPPTSQPRF
SNLSMLVLQWASLAVLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAETYV"

promoter complement(89434)
/note="TATA: TATAAAG"

CDS complement(89567. .90013)
/note="BLLF2 early reading frame (BLLF3 in Baer et al, 1984)"
/codon_start=1
/db_xref="PID:g1334867"
/db_xref="SWISS-PROT:P03199"
/translation="MCPPVRQHPAQAPPAKRQALETVPHPQNRGRLMSPKARPPKMQR
RPRPPVAKRRRFPRSPQQVERPILPPVESTPQDMEPGQVQSPPQITAVIQLRQDRDTM
RPPIYLPALLANCGPAGLLRAHRLPQPKPPCQSRQRPSPDSQTSPC"

promoter complement(90051)
/note="TATA: TATAACA BL-L2 early promoter before BLLF2, gives 0.7kb early RNA"

intron complement(90062. .90652)
/note="intervening sequence in gp220 gene"

repeat_region 90177. .90639
/note="21 copies of 21bp approximate repeat"

promoter complement(92192)
/note="TATA: TATTAAA BL-L1 late promoter before BLLF1a,b. Gives 2.8 and 2.2kb late RNAs"

mRNA 92238. .92581
/note="Exon in Bodescot et al (1986) RNA (spliced from 20763 to 92670)"

CDS 92243. .92602
/note="BLRF3 reading frame"
/codon_start=1
/db_xref="PID:g1632791"
/db_xref="SWISS-PROT:P03202"
/translation="MDKDRPGPPALDDNMEEEVPSTSVVQEQVSAGDWENVLIELSDS
SSEKEAEDAHLEPAQKGTKRKRVDHDAGGSAPARPMLPPQPDLPGREAILRRFPLDLR
TLLQAIGAAATVSIPMA"

mRNA 92670. .95248
/note="Exon in (Bodescot et al, 1986) RNA from 92581, to 3' end"

CDS 92670. .95162
/note="BERF1 frame, homology with BERF2b and BERF4. A fusion of BLRF3 with BERF1 encodes EBNA-3A, latent cycle gene. (Hennessy et al, 1986, Joab et al, 1987); Protein sequence is in conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25053"
/db_xref="PID:g1334871"
/db_xref="SWISS-PROT:P12977"
/translation="RIDTRAIDQFFGSQISNTEMYIMYAMAIRQAIRDRRRNPASRRD
QAKWRLQTLAAGWPMGYQAYSSWMYSYTDHQTTPTFVHLQATLGCTGGRRCHVTFSAG
TFKLPRCTPGDRQWLYVQSSVGNIVQSCNPRYSIFFDYMAIHRSLTKIWEEVLTPDQR
VSFMEFLGFLQRTDLSYIKSFVSDALGTTSIQTPWIDDNPSTETAQAWNAGFLRGRAY
GIDLLRTEGEHVEGATGETREESEDTESDGDDEDLPCIVSRGGPKVKRPPIFIRRLHR
LLLMRAGKRTEQGKEVLEKARGSTYGTPRPPVPKPRPEVPQSDETATSHGSAQVPEPP
TIHLAAQGMAYPLHEQHGMAPCPVAQAPPTPLPPVSPGDQLPGVFSDGRVACAPVPAP
AGPIVRPWEPSLTQAAGQAFAPVRPQHMPVEPVPVPTVALERPVYPKPVRPAPPKIAM
QGPGETSGIRRARERWRPAPWTPNPPRSPSQMSVRDRLARLRAEAQVKQASVEVQPPQ
LTQVSPQQPMEGPLVPEQQMFPGAPFSQVADVVRAPGVPAMQPQYFDLPLIQPISQGA
PVAPLRASMGPVPPVPATQPQYFDIPLTEPINQGASAAHFLPQQPMEGPLVPEQWMFP
GAALSQSVRPGVAQSQYFDLPLTQPINHGAPAAHFLHQPPMEGPWVPEQWMFQGAPPS
QGTDVVQHQLDALGYTLHGLNHPGVPVSPAVNQYHLSQAAFGLPIDEDESGEGSDTSE
PCEALDLSIHGRPCPQAPEWPVQEEGGQDATEVLDLSIHGRPRPRTPEWPVQGEGGQN

```
                VTGPETRRVVVSAVVHMCQDDEFPDLQDPPDEA"
misc_feature    92703
                /note="BAM: Bam H1 L/E"
promoter        complement(93161)
                /note="TATA: CATAAAT"
promoter        93479
                /note="TATA: TATAAGA"
promoter        complement(93482)
                /note="TATA: TATAAAT"
repeat_region   94208. .94277
                /note="repeat type A"
repeat_region   94281. .94306
                /note="repeat type B"
repeat_region   94307. .94381
                /note="repeat type C"
repeat_region   94386. .94411
                /note="repeat type B"
repeat_region   94412. .94489
                /note="repeat type C"
repeat_region   94490. .94560
                /note="repeat type A"
repeat_region   94571. .94648
                /note="repeat type C"
repeat_region   94649. .94719
                /note="repeat type A"
repeat_region   94896. .94982
                /note="repeat type D"
repeat_region   94983. .95069
                /note="repeat type D"
misc_feature    95221
                /note="polyA signal: AATAAA"
misc_feature    complement(95272)
                /note="polyA signal: AATAAA"
CDS             join(95353. .95709,95788. .98247)
                /codon_start=1
                /product="EBNA3B (EBNA4A) latent protein"
                /db_xref="PID:e276431"
                /db_xref="PID:g1632792"
                /db_xref="SWISS-PROT:P03203"
                /translation="MKKAWLSRAQQADAGGASGSEDPPDYGDQGNVTQVGSEPISPEI
                GPFELSAASEDDPQSGPVEENLDAAAREEEEPHEQEHNGGDDPLDVHTRQPRFVDVNP
                TQAPVIQLVHAVYDSMLQSDLRPLGSLFLEQNLNIEEFIWMCMTVRHRCQAIRKKPLP
                IVKQRRWKLLSSCRSWRMGYRTHNLKVNSFESGGDNVHPVLVTATLGCDEGTRHATTY
                SAGIVQIPRISDQNQKIETAFLMARRARSLSAERYTLFFDLVSSGNTLYAIWIGLGTK
                NRVSFIEFVGWLCKKDHTHIREWFRQCTGRPKAAKPWLRAHPVAIPYDDPLTNEEIDL
                AYARGQAMNIEAPRLPDDPIIVEDDDESEEIEAESDEEEDKSGMESLKNIPQTLPYNP
                TVYGRPAVFDRKSDAKSTKKCRAIVTDFSVIKAIEEEHRKKKAARTEQPRATPESQAP
                TVVLQRPPTQQEPGPVGPLSVQARLEPWQPLPGPQVTAVLLHEESMQGVQVHGSMLDL
                LEKDDEVMEQRVMATLLPPVPQQPRAGRRGPCVFTGDLGIESDEPASTEPVHDQLLPA
                PGPDPLEIQPLTSPTTSQLSSSAPSCAQTPWPVVQPSQTPDDPTKQSRPPETAAPRQW
                PMPLRPIPMRPLRMQPIPFNHPVGPTPHQTPQVEITPYKPTWAQIGHIPYQPTPTGPA
                TMLLRQWAPATMQTPPRAPTPMSPPEVPPVPRQRPRGAPTPTPPPQVPPVPRQRPRGA
                PTPTPPPQVLPTPMQLALRAPAGQQGPTKQILRQLLTGGVKKGRPSLKLQAALERQAA
                AGWQPSPGSGTSDKIVQAPIFYPPVLQPIQVMGQGGSPTAMAASAVTQAPTEYTRERR
                GVGPMPPTDIPPSKRAKIEAYTEPEMPHGGASHSPVVILENVGQGQQQTLECGGTAKQ
                ERDMLGLGDIAVSSPSSSETSNDE"
mat_peptide     95353. .95724
                /note="BERF2a reading frame"
mat_peptide     95725. .98244
                /note="BERF2b frame, homology with BERF1 and BERF4. BERF2a
                and BERF2b are spliced together to make EBNA3B (EBNA4A)
                latent protein."
misc_feature    complement(95819)
                /note="polyA signal: AATAAA"
promoter        complement(95853)
                /note="TATA: TATAAAT"
misc_feature    complement(96276)
                /note="polyA signal: AATAAA"
repeat_region   97522. .97698
                /note="3x60bp repeat"
mat_peptide     98323. .98769
                /note="BERF3 reading frame"
mRNA            98364. .98730
                /note="Exon in EBNA-1 RNA (Speck and Strominger, 1985)"
CDS             join(98371. .98730,98805. .101423)
                /codon_start=1
                /product="EBNA3C (EBNA 4B) latent protein"
                /db_xref="PID:e276455"
                /db_xref="PID:g1632793"
                /db_xref="SWISS-PROT:P03204"
                /translation="MESFEGQGDSRQSPDNERGDNVQTTGEHDQDPGPGPPSSGASER
                LVPEESYSRDQQPWGQSRGDENRGWMQRIRRRRRRRAALSGHLLDTEDNVPPWLPPHD
                ITPYTARNIRDAACRAVKQSHLQALSNLILDSGLDTQHILCFVMAARQRLQDIRRGPL
                VAEGGVGWRHWLLTSPSQSWPMGYRTATLRTLTPVPNRVGADSIMLTATFGCQNAART
                LNTFSATVWTPPHAGPREQERYAREAEVRFLRGKWQRRYRRIYDLIELCGSLHHIWQN
                LLQTEENLLDFVRFMGVMSSCNNPAVNYWFHKTIGNFKPYYPWNAPPNENPYHARRGI
                KEHVIQNAFRKAQIQGLSMLATGGEPRGDATSETSSDEDTGRQGSDVELESSDDELPY
                IDPNMEPVQQRPVMFVSRVPAKKPRKLPWPTPKTHPVKRTNVKTSDRSDKAEAQSTPE
                RPGPSEQSSVTVEPAHPTPVEMPMVILHQPPPVPKPVPVKPTPPPSRRRRGACVVYDD
                DVIEVIDVETTEDSSSVSQPNKPHRKHQDGFQRSGRRQKRAAPPTVSPSDTGPPAVGP
                PAAGPPAAGPPAAGPPAAGPPAAGPPAAGPRILAPLSAGPPAAGPHIVTPPSARPRIM
                APPVVRMFMRERQLPQSTGRKPQCFWEMRAGREITQMQQEPSSHLQSATQPTTPRPSW
                APSVCALSVMDAGKAQPIESSHLSSMSPTQPISHEEQPRYEDPDAPLDLSLHPDVAAQ
                PAPQAPYQGYQEPPAPQAPYQGYQEPPPPQAPYQGYQEPPAHGLQSSSYPGYAGPWTP
                RSQHPCYRHPWAPWSQDPVHGHTQGPWDPRAPHLPPQWDGSAGHGQDQVSQFPHLQSE
                TGPPRLQLSLVPLVSSSAPSWSSPQPRAPIRPIPTRFPPPPPMPLQDSMAVGCDSSGTA
                CPSMPFASDYSQGAFTPLDINATTPKRPRVEESSHGPARCSQATAEAQEILSDNSEIS
                VFPKDAKQTDYDASTESELD"
misc_feature    98731
                /note="DONOR: AAGGTGAGT donor"
mat_peptide     98805. .101420
                /note="BERF4 frame, homology with BERF1 and BERF2b. BERF3
                and BERF4 are spliced together to make the EBNA3C (EBNA
                4B) latent protein."
mRNA            98805. .99050
                /note="Exon in T4 cDNA (Bodescot et al 1986). 99050 is not
                the end of the RNA."
misc_feature    99126. .102118
                /note="DEL: Deletion in Raji"
promoter        99443
                /note="TATA: CATAAAA"
misc_feature    100104
                /note="DONOR: ACCGTGAGT possible donor before repeat."
repeat_region   100122. .100304
                /note="10 x 15bp repeat"
misc_feature    100528
                /note="DONOR: CTGGTAAGG possible donor"
misc_feature    100613
                /note="BAM: Bam H1 E/e1"
repeat_region   100665. .100781
                /note="3x39bp repeat"
promoter        complement(100860)
                /note="TATA: TATAACA"
misc_feature    100919
                /note="BAM: Bam H1 e1/e2"
misc_feature    101426
                /note="BAM: Bam H1 e2/e3"
CDS             complement(101445. .102116)
                /note="BZLF2 reading frame 3x NXT/S. 2.5kb late RNA
                traverses BZLF2, ends unknown."
                /codon_start=1
                /db_xref="PID:g1334876"
                /db_xref="SWISS-PROT:P03205"
                /translation="MVSFKQVRVPLFTAIALVIVLLLAYFLPPRVRGGGRVAAAAITW
                VPKPNVEVWPVDPPPPVNFNKTAEQEYGDKEVKLPHWTPTLHTFQVPQNYTKANCTYC
                NTREYTFSYKGCCFYFTKKKHTWNGCFQACAELYPCTYFYGPTPDILPVVTRNLNAIE
                SLWVGVYRVGEGNWTSLDGGTFKVYQIFGSHCTYVSKFSTVPVSHHECSFLKPCLCVS
                QRSNS"
promoter        101690
                /note="TATA: CATAAAA"
misc_feature    101765
                /note="polyA signal: AATAAA"
promoter        complement(101786)
                /note="TATA: TATAAAG"
misc_feature    101947
                /note="BAM: Bam H1 e3/Z"
```

misc_feature complement(102098)
/note="DONOR: CAGGTGAGG possible donor"
mRNA complement(102126. .102341)
/note="3' terminal exon of 0.9kb and 2.8kb early RNAs"
promoter 102153
/note="TATA: TATTAAT"
misc_feature complement(102156)
/note="polyA signal: AATAAA 3' end of 0.9kb and 2.8kb RNAs encoding BZLF1 and BRLF1"
promoter complement(102160)
/note="TATA: TATTAAT"
CDS complement(join(<102210. .102338,102423. .102530, 102655. .103155))
/note="BZLF1 reading frame, modified from Baer et al, 1984. Has two splices within frame. 2xNXT/S. Immediate early gene which disrupts latency (Countryman and Miller, 1985), called EB1 by Chevallier-Greco et al, 1986 and ZEBRA by Miller."
/codon_start=1
/db_xref="PID:e276432"
/db_xref="PID:g1632794"
/db_xref="SWISS-PROT:P03206"
/translation="MMDPNSTSEDVKFTPDPYQVPFVQAFDQATRVYQDLGGPSQAPL PCVLWPVLPEPLPQGQLTAYHVSTAPTGSWFSAPQPAPENAYQAYAAPQLFPVSDITQ NQQTNQAGGEAPQPGDNSTVQTAAAVVFACPGANQGQQLADIGVPQPAPVAAPARRTR KPQQPESLEECDSELEIKRYKNRVASRKCRAKFKQLLQHYREVAAAKSSENDRLRLLL KQMCPSLDVDSIIPRTPDVLHEDLLNF"
promoter complement(102380)
/note="TATA: CATAAAT"
promoter 102415
/note="TATA:TATATAC"
promoter complement(102420)
/note="TATA: TATATAC"
mRNA complement(102426. .102530)
/note="Exon of 0.9kb and 2.8kb early RNAs"
misc_feature complement(102504)
/note="polyA signal: AATAAA, apparently not functional"
repeat_region 102581. .102652
/note="semi-repetitive sequence, homologous to human c-fos 3' sequence"
mRNA complement(102655. .103194)
/note="First exon of 0.9kb early RNA encoding BZLF1"
misc_feature complement(102918)
/note="splice acceptor used in RZ fusion gene (Sargeant)"
promoter complement(103231)
/note="TATA: TTTAAA of BZL1 immediate early promoter gives 0.9kb RNA"
misc_feature complement(103256. .103311)
/note="Upstream of BZL1, homology to 106243 to 106188"
CDS complement(103366. .105183)
/note="BRLF1 reading frame, (immediate?) early gene, acts as transcription activator."
/codon_start=1
/db_xref="PID:g1334878"
/db_xref="SWISS-PROT:P03209"
/translation="MRPKKDGLEDFLRLTPEIKKQLGSLVSDYCNVLNKEFTAGSVEI TLRSYKICKAFINEAKAHGREWGGLMATLNICNFWAILRNNRVRRRAENAGNDACSIA CPIVMRYVLDHLIVVTDRFFIQAPSNRVMIPATIGTAMYKLLKHSRVRAYTYSKVLGV DRAAIMASGKQVVEHLNRMEKEGLLSSKFKAFCKWVFTYPVLEEMFQTMVSSKTGHLT DDVKDVRALIKTLPRASYSSHAGQRSYVSGVLPACLLSTKSKAVETPILVSGADRMDE ELMGNDGGASHTEARYSESGQFHAFTDELESLPSPTMPLKPGAQSADCGDSSSSSSDS GNSDTEQSEREEARAEAPRLRAPKSRRTSRPNRGQTPCPSNAAEPEQPWIAAVHQESD ERPIFPHPSKPTFLPPVKRKKGLRDSREGMFLPKPEAGSAISDVFEGREVCQPKRIRP FHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPVPQPLDPAPAVTPEASHLLEDP DEETSQAVKALREMADTVIPQKEEAAICGQMDLSHPPPRGHLDELTTTLESMTEDLNL DSPLTPELNEILDTFLNDECLLHAMHISTGLSIFDTSLF"
misc_feature complement(103453. .103462)
/note="TAATGAAATC sequence"
misc_feature 103741
/note="BAM: Bam H1 Z/g"
misc_feature 103816
/note="BAM: Bam H1 g/R"
mRNA complement(104926. .105185)
/note="exon in RZ fusion gene (Sargeant)"
mRNA complement(104927. .104989)
/note="BRLF2 poss. small 5' exon"
promoter 105016
/note="TATA: TATAAAT before BRRF1, possible promoter for 1.1 kb early RNA encoding BRRF1"
promoter complement(105019)
/note="TATA: TATAAAT before BRLF2"
CDS 105182. .106114
/note="BRRF1 early reading frame"
/codon_start=1
/db_xref="PID:g1334877"
/db_xref="SWISS-PROT:P03207"
/translation="MASSNRGNARPLKSFLHELYLKHYPEVGDVVHLLNTIGVDCDLP PSHPLLTAQRGLFLARVLQAVQQHKLLEDTIVPKILKKLAYFLELLSYYSPKDEQRDI AEVLDHLKTNRDLGLDDRLWALIRKLRQDRHHASVNVLMPGSDYTAVSLQYYDGISIG MRKVIADVCRSGYASMPSMTATHNLSHQLLMASGPSEEPCAWRGFFNQVLLWTVALCK FRRCIYYNYIQGSIATISQLLHLEIKALCSWIISQDGMRLFQHSRPLLTLWESVAANQ EVTDAITLPDCAEYIDLLKHTKHVLENCSAMQYK"
misc_feature complement(105185)
/note="ACCEPT: splice acceptor in 2.8kb early RNA encoding BRLF1 and RZ fusion gene (Sargeant)"
promoter 105213
/note="TATA: CATTAAA"
misc_feature 106110
/note="polyA signal: AATAAA, 3' end of early 1.1kb RNA encoding BRRF1"
misc_feature complement(106125)
/note="DONOR: CAGGTAAGA possible donor"
misc_feature complement(106188. .106243)
/note="Homology to upstream region of BZL1"
promoter complement(106213)
/note="TATA: CATAAAA"
promoter 106243
/note="TATA: TATAAAA before BRRF2, possible promoter for 1.8 kb RNA encoding BRRF2"
CDS 106302. .107915
/note="BRRF2 reading frame"
/codon_start=1
/db_xref="PID:g1334879"
/db_xref="SWISS-PROT:P03210"
/translation="MSGQQRGSVILVPEHLAGALTKLMSDFITGQDVTLSGGNIAVKI RDAINQTPGGGDVAILSSLFALWNALPTSGRQSSRDDLIPAAVQALTTAHNLCLGVIP GETSHKDTPESLLRAIVTGLQKLWVDSCGCPECLQCLKGLKAIKPGLYEIPRIIPHTK QCSPVNLLNMLVHKLVALRGHVQLAYDARVLTPDFHEIPDLDDSDAVFARTLLAALFH LNMFFILKDYITQDSMSLKQALSGHWMSATGNPLPAAPETLRDYLEAFRNSDNHFYLP TTGPLNTFQFPEELLGRVVVIDSSLCAASHVQDVITHGVGAGVPRPRFSALPPAPSRE PQQTCSQLTSRGNESSRRNLGQPGGTSPAVPPVCPIVSLTASGAKQNRGGMGSLHLAK PEETSPAVSPVCPIASPAASRSKQHCGVTGSSQAAPSFSSVAPVASLSGDLEEEEEGS RESPSLPSSKKGDEEFEAWLEAQDANLEDVQREFSGLRVIGDEDEDGSEDGEFSDLDL SDSDHEGDEGGGAVGGGRSLHSLYSLSVV"
promoter complement(106385)
/note="TATA: GATAAAA"
misc_feature complement(106973)
/note="polyA signal: AATAAA"
promoter complement(107124)
/note="TATA: GATAAAA"
misc_feature 107457
/note="BAM: Bam H1 R/f"
misc_feature 107565
/note="BAM: Bam H1 f/K"
misc_feature 107914
/note="polyA signal: AATAAA, 3' end of 1.8kb RNA encoding BBRF2"
misc_feature 107942
/note="ACCEPT: splice acceptor for EBNA-1 RNA (from 98730)"
CDS 107950. .109875
/note="BKRF1 encodes EBNA-1 protein, latent cycle gene."
/codon_start=1
/db_xref="PID:g1334880"
/db_xref="SWISS-PROT:P03211"
/translation="MSDEGPGTGPGNGLGEKGDTSGPEGSGGSGPQRRGGDNHGRGRG

RGRGRGGGRPGAPGGSGSGPRHRDGVRRPQKRPSCIGCKGTHGGTGAGAGAGGAGAGG
AGAGGGAGAGGGAGGAGGAGGAGAGGGAGAGGGAGGAGGAGAGGGAGGAGGAGAG
GGAGGAGGAGAGGGAGAGGGAGGAGAGGGAGGAGGAGAGGGAGAGGAGGAGGAGA
GAGGGAGGAGGAGAGGAGAGGAGAGGAGAGGAGGAGAGGAGGAGAGGGAGAGGAG
GAGAGGGAGGAGAGGAGGAGAGGAGGAGAGGAGGAGAGGGAGAGGAGAGGGRGGS
GGRGRGGSGGRGRGGSGGRRGRGRERARGGSRERARGRGRGRGEKRPRSPSSQSSSSG
SPPRRPPPGRRPFFHPVGEADYFEYHQEGGPDGEPDVPPGAIEQGPADDPGEGPSTGP
RGQGDGGRRKKGGWFGKHRGQGGSNPKFENIAEGLRALLARSHVERTTDEGTWVAGVF
VYGGSKTSLYNLRRGTALAIPQCRLTPLSRLPFGMAPGPGPQPGPLRESIVCYFMVFL
QTHIFAEVLKDAIKDLVMTKPAPTCNIRVTVCSFDDGVDLPPWFPPMVEGAAAEGDDG
DDGDEGGDGDEGEEGQE"

repeat_region 108217. .108924
/note="EBNA triplet repeat GGA,GCA,GGG."

misc_feature 109856
/note="DONOR: AGGGTGAGG possible donor at end BKRF1"

promoter 109905
/note="TATA: TATTAAA before BKRF2, possible start for 2.3kb late RNA"

misc_feature 109906
/note="polyA signal: ATTAAA"

misc_feature 109937
/note="polyA signal: AATAAA 3' end of EBNA-1 RNA"

CDS 109958. .110371
/note="BKRF2 reading frame"
/codon_start=1
/db_xref="PID:g1334881"
/db_xref="SWISS-PROT:P03212"
/translation="MRAVGVFLAICLVTIFVLPTWGNWAYPCCHVTQLRAQHLLALEN
ISDIYLVSNQTCDGFSLASLNSPKNGSNQLVISRCANGLNVVSFFISILKRSSSALTG
HLRELLTTLETLYGSFSVEDLFGANLNRYAWHRGG"

misc_feature 110271
/note="DONOR: TCCGTGAGT possible donor at end of BKRF2"

CDS 110353. .111120
/note="BKRF3 reading frame, homologous to RF 59 VZV"
/codon_start=1
/db_xref="PID:e25020"
/db_xref="PID:g1632795"
/db_xref="SWISS-PROT:P12888"
/translation="MASRGLDLWLDEHVWKRKQEIGVKGENLLLPDLWLDFLQLSPIF
QRKLAAVIACVRRLRTQATVYPEEDMCMAWARFCDPSDIKVVILGQDPYHGGQANGLA
FSVAYGFPVPPSLRNIYAELHRSLPEFSPPDHGCLDAWASQGVLLLNTILTVQKGKPG
SHADIGWAWFTDHVISLLSERLKACVFMLWGAKAGDKASLINSKKHLVLTSQHPSPLA
QNSTRKSAQQKFLGNNHFVLANNFLREKGLGEIDWRL"

misc_feature 111098
/note="DONOR: TCGGTGAGA possible donor at end BKRF3"

CDS 111134. .111787
/note="BKRF4 reading frame, contains complex repetitive sequence"
/codon_start=1
/db_xref="PID:e25021"
/db_xref="PID:g1632796"
/db_xref="SWISS-PROT:P30117"
/translation="MAMFLKSRGVRSCRDRRLLSDEEEETSQSSSYTLGSQASQSIQE
EDVSDTDESDYSDEDEEIDLEEEYPSDEDPSEGSDSDPSWHPSDSDESDYSESDEDEA
TPGSQASRSSRVSPSTQQSSGLTPTPSFSRPRTRAPPRPPAPAPVRGRASAPPRPPAP
VQQSTKDKGPHRPTRPVLRGPAPRRPPPPSSPNTYNKHMMETTPPIKGNNNYNWPWL"

misc_feature 111272
/note="DONOR: GACGTGAGT poss.donor before rpt.seq. in BKRF4"

misc_feature 111719
/note="polyA signal: AATAAA"

misc_feature 111787
/note="polyA signal: AATAAA : currently unknown which is 3' end of the 2.3kb late and 1.1kb early RNAs"

CDS complement(111830. .114259)
/note="BBLF4 early reading frame, very good homology to RF55 VZV"
/codon_start=1
/db_xref="PID:g1334885"
/db_xref="SWISS-PROT:P03214"
/translation="MAEEPRAPEALSSTFMLNMTSDASVRRIVRRIGTLARRRVQQLP
DMETFSPEFDPELSEPPFLPFSAYVITGTAGAGKSTSVSCLHHTMDCLVTGATTVAAQ
NLSQTLRAYCPTVYSAFGFKSRHINMTQRVSSHGRSTDAALEELQRRDLAKYWPVLSD
IAAEFRRTKPRGLYSGVSGPAFEVLRDMHQGQLWTTNVIVVDEAGTLSVHILTAVVFC
YWFFNAWLRTPLYRRGRIPCIVCVGSPTQTDAFQSSFSHETQVNKIRECDNILTFLVG
NPRAATYVDVARNWALFINNKRCTDVQFGHLMKTLEYGLELSPDILAYVDRFVVPRAA
IMDPAQYVGWTRLFLSHAEVKTFLTTLHATLKTAGQGRAARGTGGDGGGVTMFTCPVE
CEVFLDPLAQYKTLVGLPGLTAHTWLQKNYARLGNYSQFADQDMVPVGTEQDEERVKV
TYNVTYVKHSSVSVNCKTKKSICGYTGTFGDFMDTLEADSFVEAHGHEQPEYVYSFLA
RLIYGGIYAFSHGGHSLCENGEYVAELGAVPLPGRTWDPEVTAGMELGELPLEVAWDG
ERSPAAVFYARVLAPPAANSAPLCSLLNIYNDLRAYFRQCLDVAVRYGGREFRDLPFC
TFTNNMLIRDNIEFTSDEPLLHGLLDYASTTENYTLLGYTHLNVFFGIRGKQQPQDAG
SSRMPRLMVKDEAGFVCCLEHNTNKLYETIEDKSLNLCSIRDYGISSKLAMTIAKAQG
LSLNKVAICFGSHRNIKPGHVYVALSRARHSNCVVMDRNPLSEMITGEGNPASGYIVD
ALKNSRALLVY"

misc_feature complement(111830)
/note="polyA signal: AATAAA"

promoter 112471
/note="TATA: TATATAT"

promoter complement(112476)
/note="TATA: TATATAA"

misc_feature 112620
/note="BAM: Bam H1 K/B"

promoter 113876
/note="TATA: TATTTAT before BBRF1"

promoter complement(113885)
/note="TATA: CATAAAT"

CDS 114204. .116045
/note="BBRF1 late reading frame, homologous to RF 54 VZV"
/codon_start=1
/db_xref="PID:g1334884"
/db_xref="SWISS-PROT:P03213"
/translation="MFNMNVDESASGALGSSAIPVHPTPASVRLFEILQGKYAYVQGQ
TIYANLRNPGVFSRQVFTHLFKRAISHCTYDDVLHDWNKFEACIQKRWPSDDSCASRF
RESTFESWSTTMKLTVRDLLTTNIYRVLHSRSVLSYERYVDWICATGMVPAVKKPITQ
ELHSKIKSLRDRCVCRELGHERTIRSIGTELYEATKEIIESLNSTFIPQFTEVTIEYL
PRSDEYVAYYCGRRIRLHVLFPPAIFAGTVTFDSPVQRLYQNIFMCYRTLEHAKICQL
LNTAPLKAIVGHGGRDMYKDILAHLEQNSQRKDPKKELLNLLVKLSENKTISGVTDVV
EEFITDASNNLVDRNRLFGQPGETAAQGLKKKVSNTVVKCLTDQINEQFDQINGLEKE
RELYLKKIRSMESQLQASLGPGGNNPAASAPAAVAAEAASVDILTGSTASAIEKLFNS
PSASLGARVSGHNESILNSFVSQYIPPSREMTKDLTELWESELFNTFKLTPVVDNQGQ
RLYVRYSSDTISILLGPFTYLVAELSPVELVTDVYATLGIVEIIDELYRSSRLAIYIE
DLGRKYCPASATGGDHGIRQAPSARGDTEPDHAKSKPARDPPPGAGS"

CDS 115948. .116784
/note="BBRF2 late reading frame, homologous to RF 53 VZV"
/codon_start=1
/db_xref="PID:e25024"
/db_xref="PID:g1632797"
/db_xref="SWISS-PROT:P29882"
/translation="MASGKHHQPGGTRSLTMQKVSLRVTPRLVLEVNRHNAICVATNV
PEFYNARGDLNIRDLRAHVKARMISSQFCGYVLVSLLDSEDQVDHLNIFPHVFSERMI
LYKPNNVNLMEMCALLSMIENAKSPSIGLCREVLGRLTLLHSKCNNLDSLFLYNGART
LLSTLVKYHDLEEGAATPGPWNEGLSLFKLHKELKRAPSEARDLMQSLFLTSGKMGCL
ARSPKDYCADLNKEEDANSGFTFNLFYQDSLLTKHFQCQTVLQTLRRKCLGSDTVSKI
IP"

promoter complement(116683)
/note="TATA: GATAAAA"

misc_feature complement(116696)
/note="polyA signal: AATAAA"

CDS complement(116784. .117386)
/note="BBLF3 early reading frame, spliced to BBLF2. BBLF3 contains a consensus nucleotide binding site; Protein sequence is in conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25025"
/db_xref="PID:g1334887"
/db_xref="SPTREMBL:Q04343"
/translation="AFLQGVKDSEDASRLDRDVMGGEATVARRHIRVKARRGPGCLLM
AIFQGDLYVGGCREHSGPFLVWHEAFSWTLDQLAARPEADKAPPSHDHLLTLVRDLTR
RLAPGRRRNRFWALPRAWLQRLRRAGLRLSGSHVCLLDKDGARPAPCQTATEHGLSPT
AYFREIMAFLLDVISALHPGYTIPMEITRETDLLMTVLSLF"

misc_feature 116785
/note="polyA signal: AATAAA"

intron complement(117386. .117515)
/note="intron spliced out in RNA linking BBLF2 and BBLF3"

CDS complement(117515. .119080)

```
                /note="BBLF2 early reading frame, spliced to BBLF3;
                Protein sequence is in conflict with the conceptual
                translation"
                /codon_start=1
                /db_xref="PID:e25026"
                /db_xref="PID:g1334888"
                /db_xref="SWISS-PROT:P30118"
                /translation="PDVLKGPVLLRSQTMMETPAESVRARVSSVTFYNVTQTAGRWWA
                IWVVGIVPIKREDVETLIVVQACQPPLGGSLEPPVVNAPSTTELNFLRWERELRRSGG
                LIAMLADAAEKDLFDLSFRTRDRRLLSAARVEDEQGLIFQPLFPAQVVCQSCSGDDGR
                DQQPPPVDGFGSEMEGEQTCPHAQRHSESPGQLDVYIRTPRGDVFTYSTETPDDPSPV
                PFRDILRPVTYEVDLVSSDGATGRGGDARRHRVSLKILEPAGGFESWLVNSWSMAGGG
                LYAFLRSIYASCYANHRGTKPIFYLLDPELCPGGSDFQPYVPGFPFLPIHYVGRARPA
                FWHRAPHSEGLLLLDLNLGVSGTPLADALLGLDARSGQRRGSLLLQQIWPPTRKEINP
                RHVCTREGGEGGGEDETTVVGRAEATAILEADATWWLYELARCHLSARGAPVGTPDGG
                GQARDAQTWLRALHRYGTSDTRRALGGLYTAVTRVLLHAAADLGLTWAYADEFILGFV
                APTSAHPSEEPLAQ"
promoter        118981
                /note="TATA: TATAAAA BBR1 late promoter before BBRF3"
promoter        119067
                /note="TATA: TTTAAAA BBR2 late promoter ?"
promoter        119098
                /note="TATA: TATTTAA BBR3 late promoter before BBRF3"
misc_feature    119108
                /note="DONOR: AAGGTGAAT possible donor"
CDS             119137. .120354
                /note="BBRF3 late reading frame"
                /codon_start=1
                /db_xref="PID:g1334889"
                /db_xref="SWISS-PROT:P03215"
                /translation="MKSSKNDTFVYRTWVKTLVVYFVMFVMSAVVPITAMFPNLGYPC
                YFNALVDYGALNLTNYNLAHHLTPTLYLEPPEMFVYITLVFIADCVAFIYYACGEVAL
                IKARKKVSGLTDLSAWVSAVGSPTVLFLAILKLWSIQVFIQVLSYKHVFLSAFVYFLH
                FLASVLHACACVTRFSPVWVVKAQDNSIPQDTFLWWVVFYLKPVVTNLYLGCLALETL
                VFSLSVFLALGNSFYFMVGDMVLGAVNLFLILPIFWYILTEVWLASFLRHNFGFYCGM
                FIASIILILPLVRYEAVFVSAKLHTTVAINVAIIPILCSVAMLIRICRIFKSMRQGTD
                YVPVSETVELELESEPRPRPSRTPSPGRNRRRSSTSSSSSRSTRRQRPVSTQALVSSV
                LPMTTDSEEEIFP"
misc_feature    120260
                /note="ACCEPT: ATCTTCCTCCAGGT possible acceptor"
misc_feature    120358
                /note="polyA signal: AATAAA"
CDS             complement(120747. .120974)
                /note="BBLF1 late reading frame, possibly homologous to RF
                49 VZV"
                /codon_start=1
                /db_xref="PID:g1334890"
                /db_xref="SWISS-PROT:P03216"
                /translation="MGALWSLCRRRVNSIGDVDGGIINLYNDYEEFNLETTKLIAAEE
                GRACGETNEGLEYDEDSENDELLFLPNKKPN"
misc_feature    complement(120764)
                /note="polyA signal: AATAAA, 3' end of 0.6kb late, 1.6kb
                early, 3.0kb early RNAs"
CDS             complement(120929. .122341)
                /note="BGLF5 early reading frame, homologous to RF 48 VZV
                and alkaline exonuclease of HSV"
                /codon_start=1
                /db_xref="PID:g1334891"
                /db_xref="SWISS-PROT:P03217"
                /translation="MADVDELEDPMEEMTSYTFARFLRSPETEAFVRNLDRPPQMPAM
                RFVYLYCLCKQIQEFSGETGFCDFVSSLVQENDSKDGPSLKSIYWGLQEATDEQRTVL
                CSYVESMTRGQSENLMWDILRNGIISSSKLLSTIKNGPTKVFEPAPISTNHYFGGPVA
                FGLRCEDTVKDIVCKLICGDASANRQFGFMISPTDGIFGVSLDLCVNVESQGDFILFT
                DRSCIYEIKCRFKYLFSKSEFDPIYPSYTALYKRPCKRSFIRFINSIARPTVEYVPDG
                RLPSEGDYLLTQDEAWNLKDVRKRKLGPGHDLVADSLAAANRGVESMLYVMTDPSENAG
                RIGIKDRVPVNIFINPRHNYFYQVLLQYKIVGDYVRHSGGGKPGRDCSPRVNIVTAFF
                RKRSPLDPATCTLGSDLLLDASVEIPVAVLVTPVVLPDSVIRKTLSTAAGSWKAYADN
                TFDTAPWVPSGLFADDESTP"
promoter        complement(121331)
                /note="TATA: TATTAAA BBL1 late promoter before BBLF1"
promoter        121669
                /note="TATA: CATAAAT"
promoter        121697
                /note="TATA: TATAAAG"
promoter        121772
                /note="TATA: CATAAAG"
misc_feature    122313
                /note="BAM: Bam H1 B/G"
CDS             complement(122328. .123692)
                /note="BGLF4 early reading frame, homologous to RF 47 VZV;
                Protein sequence is in conflict with the conceptual
                translation"
                /codon_start=1
                /db_xref="PID:e25030"
                /db_xref="PID:g1334892"
                /db_xref="SWISS-PROT:P13288"
                /translation="SGWRSSVSRSLRPETSCDRPSSHLRNMDVNMAAELSPTNSSSSG
                ELSVSPEPPRETQAFLGKVTVIDYFTFQHKHLKVTNIDDMTETLYVKLPENMTRCDHL
                PITCEYLLGRGSYGAVYAHADNATVKLYDSVTELYHELMVCDMIQIGKATAEDGQDKA
                LVDYLSACTSCHALFMPQFRCSLQDYGHWHDGSIEPLVRGFQGLKDAVYFLNRHCGLF
                HSDISPSNILVDFTDTMWGMGRLVLTDYGTASLHDRNKMLDVRLKSSKGRQLYRLYCQ
                REPFSIAKDTYKPLCLLSKCYILRGAGHIPDPSACGPVGAQTALRLDLQSLGYSLLYG
                IMHLADSTHKIPYPNPDMGFDRSDPLYFLQFAAPKVVLLEVLSQMWNLNLDMGLTSCG
                ESPCVDVTAEHMSQFLQWCRSLKKRFKESYFFNCRPRFEHPHLPGLVAELLADDFFGP
                DGRRG"
misc_feature    complement(123506)
                /note="DONOR: AAGGTGACT possible donor"
CDS             complement(123941. .124939)
                /note="BGLF3 reading frame"
                /codon_start=1
                /db_xref="PID:g1334894"
                /db_xref="SWISS-PROT:P03220"
                /translation="MFNAVKADMPDDPMLARRYGQCLELALEACQDTPEQFKLVETPL
                KSFLLVSNILPQDNRPWHEARSSGRVAEDDYDFSSLALELLPLNPRLPEEWQFGGQGW
                SSRMEPSQPEMGMGLCFEVFDGDLMRIALAWNKDEVIGQALQILAHSQTWTSLVPEDP
                LPWMWALFYGPRSHCEERHCVYAAARGKRGPILLPTAVYTPCANIEAFLAHLTRCVYA
                LYLDVRDWKGEDIAPPFDVSRLNKMAKQLCLLPQEPFCITRVCLLCLLHKQNLNAQYK
                RPVDTYDPCLILTGEAERYMVDAVGNYREASTGTTVLYPTYDLGSIVADMVTYEDE"
promoter        complement(124117)
                /note="TATA: TATAAAA"
misc_feature    complement(124219)
                /note="polyA signal: AATAAA"
CDS             124938. .125915
                /note="BGRF1 reading frame, homologous to RF 45 VZV and
                spliced HSV gene (Costa et al, 1985). Spliced to BDRF1.
                Northern blots in BGRF1 detect 2.7, 2.6, 2.1kb late and
                1.9kb early RNAs. 2.6, 2.1kb RNAs very weak."
                /codon_start=1
                /db_xref="PID:g1334893"
                /db_xref="SWISS-PROT:P03219"
                /translation="MLYASQRGRLTENLRNALQQDSTTQGCLGAETPSIMYTGAKSDR
                WAHPLVGTIHASNLYCPMLRAYCRHYGPRPVFVASDESLPMFGASPALHTPVQVQMCL
                LPELRDTLQRLLPPPNLEDSEALTEFKTSVSSARAILEDPNFLEMREFVTSLASFLSG
                QYKHKPARLEAFQKQVVLHSFYFLISIKSLEITDTMFDIFQSAFGLEEMTLEKLHIFK
                QKASVFLIPRRHGKTWIVVAIISLILSNLSNVQIGYVAHQKHVASAVFTEIIDTLTKS
                FDSKRVEVNKETSTITFRHSGKISSTVMCATCFNKNVRPDVSVLGNCRA"
CDS             join(124938. .125873,129215. .130351)
                /note="BDRF1 reading frame"
                /codon_start=1
                /product="probable DNA packaging protein"
                /db_xref="PID:e276433"
                /db_xref="PID:g1632798"
                /db_xref="SWISS-PROT:P03219"
                /translation="MLYASQRGRLTENLRNALQQDSTTQGCLGAETPSIMYTGAKSDR
                WAHPLVGTIHASNLYCPMLRAYCRHYGPRPVFVASDESLPMFGASPALHTPVQVQMCL
                LPELRDTLQRLLPPPNLEDSEALTEFKTSVSSARAILEDPNFLEMREFVTSLASFLSG
                QYKHKPARLEAFQKQVVLHSFYFLISIKSLEITDTMFDIFQSAFGLEEMTLEKLHIFK
                QKASVFLIPRRHGKTWIVVAIISLILSNLSNVQIGYVAHQKHVASAVFTEIIDTLTKS
                FDSKRVEVNKETSTITFRHSGKISSTVMCATCFNKNSIRGQTFHLLFVDEANFIKKEA
                LPAILGFMLQKDAKIIFISSVNSADQATSFLYKLKDAQERLLNVVSYVCQEHRQDFDM
                QDSMVSCPCFRLHIPSYITMDSNIRATTNLFLDGAFSTELMGDTSSLSQGSLSRTVRD
                DAINQLELCRVDTLNPRVAGRLASSLYVYVDPAYTNNTSASGTGIAAVTHDRADPNRV
                IVLGLEHFFLKDLTGDAALQIATCVVALVSSIVTLHPHLEEVKVAVEGNSSQDSAVAI
                ASIIGESCPLPCAFVHTKDKTSSLQWPMYLLTNEKSKAFERLIYAVNTASLSASQVTV
                SNTIQLSFDPVLYLISQIRAIKPIPLRDGTYTYTGKQRNLSDDVLVALVMAHFLATTQ
                KHTFKKVH"
```

```
promoter        complement(125113)
                /note="TATA: TATAAAT before BGLF3"
misc_feature    complement(125484)
                /note="polyA signal: AATAAA, 3' end of 1.6kb late, 1.8kb
                late, 3.0kb late and 3.7kb early RNAs"
CDS             complement(125863. .126873)
                /note="BGLF2 late reading frame, poor homology to RF 44
                VZV"
                /codon_start=1
                /db_xref="PID:g1334895"
                /db_xref="SWISS-PROT:P03221"
                /translation="MASAANSSREQLRKFLNKECLWVLSDASTPQMKVYTATTAVSAV
                YVPQIAGPPKTYMNVTLIVLKPKKKPTYVTVYINGTLATVARPEVLFTKAVQGPHSLT
                LMYFGVFSDAVGEAVPVEIRGNPVVTCTDLTTAHVFTTSTAVKTVEELQDITPSEIIP
                LGRGGAWYAEGALYMFFVNMDMLMCCPNMPTFPSLTHFINLLTRCDNGECVTCYGAGA
                HVNILRGWTEDDSPGTSGTCPCLLPCTALNNDYVPITGHRALLGLMFKPEDAPFVVGL
                RFNPPKMHPDMSRVLQGVLANGKEVPCTAQPWTLLRFSDLYSRAMLYNCQVLKRQVLH
                SY"
promoter        126277
                /note="TATA: GATAAAA"
CDS             complement(126851. .128374)
                /note="BGLF1 late reading frame"
                /codon_start=1
                /db_xref="PID:g1334896"
                /db_xref="SWISS-PROT:P03222"
                /translation="MDVHIDNQVLSGLGTPLLVHLFVPDTVMAELCPNRVPNCEGAWC
                QTLFSDRTGLTRVCRVFAARGMLPGRPSHRGTFTSVPVYCDEGLPELYNPFHVAALRF
                YDEGGLVGELQIYYLSLFEGAKRALTDGHLIREASGVQESAAAMQPIPIDPGPPGGAG
                IEHMPVAAAQVEHPKTYDLKQILLEITQEENRGEQRLGHAGSPALCLGLRLRAGAETK
                AAAETSVSKHHPALENPSNIRGSAGGEGGGGRAGTGGTVGVGSGALSRVPVSFSKTRR
                AIRESRALVRGIAHIFSPHALYVVTYPELSAQGRLHRMTAVTHASPATDLAEVSILGA
                PEREFRFLISVALRISASFREKLAMQAWTAQQEIPVVIPTSYSRIYKNSDLIREAFFT
                VQTRVSWESCWVKATISNAPKTPDACLWIDSHPLYEEGASAWGKVIDSRPPGGLVGAA
                SQLVALGTDGHCVHLATTSDGQAFLVLPGGFVIKGQLALTPEERGYILARHGIRREQ"
promoter        complement(126929)
                /note="TATA: TATTAAA EEL8 late promoter before BGLF2,
                gives 1.6kb late RNA"
promoter        complement(127237)
                /note="TATA: TATAAAA, potential promoter for 1.8kb late
                RNA"
misc_feature    128029
                /note="polyA signal: AATAAA"
CDS             complement(128344. .129021)
                /note="BDLF4 early reading frame"
                /codon_start=1
                /db_xref="PID:g1334897"
                /db_xref="SWISS-PROT:P03223"
                /translation="MSDQGRLSLPRGEGGTDEPNPRHLCSYSKLEFHLPLPESMASVF
                ACWGCGEYHVCDGSSECTLIETHEGVVCALTGNYMGPHFQPALRPWTEIRQDTQDQRD
                KWEPEQVQGLVKTVVNHLYHYFLNENVISGVSEALFDQEGALRPHIPALVSFVFPCCL
                MLFRGASSEKVVDVVLSLYIHVIISIYSQKTVYGALLFKSTRNKRYDAVAKRMRELWM
                STLTTKC"
promoter        complement(128432)
                /note="TATA: TATTTAA before BGLF1, potential promoter for
                3.0kb late RNA"
misc_feature    128848
                /note="BAM: Bam H1 G/D"
promoter        complement(129054)
                /note="TATA: TATTTGC before BDLF4, potential promoter for
                3.7kb early RNA"
mat_peptide     129188. .130348
                /note="BDRF1 reading frame, homologous to RF 42 VZV and
                spliced gene in HSV (Costa et al, 1985). Spliced from
                BGRF1. Northern blots in BDRF1 detect 2.7, 2.6 kb late and
                1.9kb early RNAs. Possibly also 1.8kb early RNA."
promoter        129374
                /note="TATA: TATAAGC"
promoter        complement(129377)
                /note="TATA: TATAAAG"
misc_feature    129413
                /note="DONOR: GTGGTAAGT possible donor"
misc_feature    130347
                /note="polyA signal: ATTAAA"
misc_feature    complement(130359)
                /note="polyA signal: AATAAA, 3' end of 0.9kb late RNA,
                2.3kb late RNA and 3.2kb late RNA"
CDS             complement(130362. .131066)
                /note="BDLF3 late reading frame 9xNXT/S"
                /codon_start=1
                /db_xref="PID:g1334899"
                /db_xref="SWISS-PROT:P03224"
                /translation="MAHARDKAGAVMAMILICETSLIWTSSGSSTASAGNVTGTTAVT
                TPSPSASGPSTNQSTTLTTTSAPITTTAILSTNTTTVTFTGTTVTPVPTTSNASTINV
                TTKVTAQNITATEAGTGTSTGVTSNVTTRSSSTTSATTRITNATTLAPTLSSKGTSNA
                TKTTAELPTVPDERQPSLSYGLPLWTLVFVGLTFLMLILIFAAGLMMSAKNKPLDEAL
                LTNAVTRDPSLYKGLV"
promoter        complement(131104)
                /note="TATA: TATAAAA EEL4 late promoter before BDLF3,
                gives 0.9kb late RNA"
CDS             complement(131127. .132389)
                /note="BDLF2 late reading frame"
                /codon_start=1
                /db_xref="PID:g1334900"
                /db_xref="SWISS-PROT:P03225"
                /translation="MVDEQVAVEHGTVSHTISREEDGVVHERRVLASGERVEVFYKAP
                APRPREGRASTFHDFTVPAAAAVPGPEPEPEPHPPPMPIHANGGGETKTNTQDQNQNQT
                TRTRTNAKAEERTAEMDDTMASSGGQRGAPISADLLSLSSLTGRMAAMAPSWMKSEVC
                GERMRFKEDVYDGEAETLAEPPRCFMLSFVFIYYCCYLAFLALLAFGFNPLFLPSFMP
                VGAKVLRGKGRDFGVPLSYGCPTNPFCKVYTLIPAVVINNVTYYPNNTDSHGGHGGFE
                AAALHVAALFESGCPNLQAVTNRNRTFNVTRASGRVERRLVQDMQRVLASAVVVMHHH
                CHYETYYVFDGVGPEEFGTIPTPCFKDVLAFRPSLVTNCTAPLKTSVKGPNWSGAAGGM
                KRKQCRVDRLTDRSFPAYLEEVMYVMVQ"
promoter        132266
                /note="TATA: TATAAAA"
CDS             complement(132400. .133305)
                /note="BDLF1 late reading frame, poor homology to RF 41
                VZV"
                /codon_start=1
                /db_xref="PID:g1334901"
                /db_xref="SWISS-PROT:P25214"
                /translation="MDLKVVVSLSSRLYTDEIAKMQQRIGCILPLASTHGTQNVQGLG
                LGQVYSLETVPDYVSMYNYLSDCTLAVLDEVSVDSLILTKIVPGQTYAIKNKYQPFFQ
                WHGTGSLSVMPPVFGREHATVKLESNDVDIVFPMVLPTPIAEEVLQKILLFNVYSRVV
                MQAPGNADMLDVHMHLGSVSYLGHHYELALPEVPGPLGLALLDNLSLYFCIMVTLLPR
                ASMRLVRGLIRHEHHDLLNLFQEMVPDEIARIDLDDLSVADDLSRMRVMMTYLQSLAS
                LFNLGPRLATAAYSQETLTATCWLR"
promoter        complement(132476)
                /note="TATA: TATTTAA before BDLF2, likely promoter for
                2.3kb late RNA"
misc_feature    complement(133312)
                /note="polyA signal: AATAAA, 3' end of 4.5kb late RNA"
CDS             complement(133321. .137466)
                /note="BcLF1 late reading frame, homologous to RF 40 VZV
                and major capsid protein of HSV"
                /codon_start=1
                /db_xref="PID:g1334902"
                /db_xref="SWISS-PROT:P03226"
                /translation="MASNEGVENRPFPYLTVDADLLSNLRQSAAEGLFHSFDLLVGKD
                AREAGIKFEVLLGVYTNAIQYVRFLETALAVSCVNTEFKDLSRMTDGKIQFRISVPTI
                AHGDGRRPSKQRTFIVVKNCHKHHISTEMELSMLDLEILHSIPETPVEYAEYVGAVKT
                VASALQFGVDALERGLINTVLSVKLRHAPPMFILQTLADPTFTERGFSKTVKSDLIAM
                FKRHLLEHSFFLDRAENMGSGFSQYVRSRLSEMVAAVSGESVLKGVSTYTTAKGGEPV
                GGVFIVTDNVLRQLLTFLGEEADNQIMGPSSYASFVVRGENLVTAVSYGRVMRTFEHF
                MARIVDSPEKAGSTKSDLPAVAAGVEDQPRVPISAAVIKLGNHAVAVESLQKMYNDTQ
                SPYPLNRRMQYSYYFPVGLFMPNPKYTTSAAIKMLDNPTQQLPVEAWIVNKNNLLLAF
                NLQNALKVLCHPRLHTPAHTLNSLNAAPAPRDRRETYSLQHRRPNHMNVLVIVDEFYD
                NKYAAPVTDIALKCGLPTEDFLHPSNYDLLRLELHPLYDIYIGRDAGERARHRAVHRL
                MVGNLPTPLAPAAFQEARGQQFETATSLAHVVDQAVIETVQDTAYDTAYPAFFYVVEA
                MIHGFEEKFVMNVPLVSLCINTYWERSGRLAFVNSFSMIKFICRHLGNNAISKEAYSM
                YRKIYGELIALEQALMRLAGSDVVGDESVGQYVCALLDPNLLPPVAYTDIFTHLLTVS
                DRAPQIIIGNEVYADTLAAPQFIERVGNMDEMAAQFVALYGYRVNGDHDHDFRLHLGP
                YVDEGHADVLEKIFYYVFLPTCTNAHMCGLGVDFQHVAQTLAYNGPAFSHHFTRDEDI
                LDNLENGTLRDLLEISDLRPTVGMIRDLSASFMTCPTFTRAVRVSVDNDVTQQLAPNP
                ADKRTEQTVLVNGLVAFAFSERTRAVTQCLFHAIPFHMFYGDPRVAATMHQDVATFVM
                RNPQQRAVEAFNRPEQLFAEYREWHRSPMGKYAAECLPSLVSISGMTAMHIKMSPMAY
                IAQAKLKIHPGVAMTVVRTDEILSENILFSSRASTSMFIGTPNVSRREARVDAVTFEV
```

```
                     HHEMASIDTGLSYSSTMTPARVAAITTDMGIHTQDFFSVFPAEAFGNQQVNDYIKAKV
                     GAQRNGTLLRDPRTYLAGMTNVNGAPGLCHGQQATCEIIVTPVTADVAYFQKSNSPRG
                     RAACVVSCENYNQEVAEGLIYDHSRPDAAYEYRSTVNPWASQLGSLGDIMYNSSYRQT
                     AVPGLYSPCRAFFNKEELLRNNRGLYNMVNEYSQRLGGHPATSNTEVQFVVIAGTDVF
                     LEQPCSFLQEAFPALSASSRALIDEFMSVKQTHAPIHYGHYIIEEVAPVRRILKFGNK
                     VVF"
misc_feature         complement(133332)
                     /note="DONOR: AAGGTGGTT possible donor"
promoter             complement(133352)
                     /note="TATA: TATTAAA before BDLF1"
promoter             complement(133386)
                     /note="TATA: TATATAA"
misc_feature         135178
                     /note="polyA signal: AATAAA"
promoter             135394
                     /note="TATA: TATAAGT"
misc_feature         136624
                     /note="polyA signal: AATAAA"
misc_feature         136868
                     /note="BAM: Bam H1 D/c"
promoter             complement(137710)
                     /note="TATA: TATTAAA EHL1 promoter before BcLF1, gives
                     4.5kb late RNA"
promoter             137857
                     /note="TATA: CATAAAC"
CDS                  137991. .139718
                     /note="BcRF1 reading frame"
                     /codon_start=1
                     /db_xref="PID:e25010"
                     /db_xref="PID:g1632799"
                     /db_xref="SWISS-PROT:P25215"
                     /translation="MLAHLNQVTRIPPCPPFSGREARLKFHFFSWSTFMLSWPNNATL
                     REIRTRAATNLTHHPHLVDTLYHASPQTPFLTRSGALYRFVTCCNCTLPNISIQQCKA
                     GDRPGDLEIILQSNGGGRPASFQFPSSPTGSLLRCIVAASLLPEVSVGHQELSPLRSR
                     SQGGQTDVRSGPDPARRLVALLRREDGAPKDPPLGPFGHPRGPGPAKSEDEESERRDA
                     PPPPLDSSFQASRLVPVGPGFRLLVFNTNRVINTKLVCSEPLVKMRVCNVPRLINNFV
                     ARKYVVKETAFTVSLFFTDGVGANLAINVNISGTYLSFLLAMTSLRCFLPVEAIYPAA
                     VSNWNSTLDLHGLENQSLVRENRSGVFWTTNFPSVVSCRDGLNVSWFKAATATISRVH
                     GQTLEQHLIREITPIVTHREAKISRIKNRLFTLLELRNRSQIQVLHKRFLEGLLDCAS
                     LLRLDPSCINRIASEGLFDFSKRSIAHSKNRHECALLGHRHSANVTKLVVNERKTRLD
                     ILGRNANFLTRCKHQVNLRQSPIFLTLLRHIRRRLGLGRASVKREITLLLAHLRKKTA
                     PIHCRDAQV"
misc_feature         138019
                     /note="BAM: Bam H1 c/b"
misc_feature         139352
                     /note="BAM: Bam H1 b/T"
CDS                  139642. .140916
                     /note="BTRF1 reading frame. Northern blots detect 0.95
                     late and 3.8kb early RNA; Protein sequence is in conflict
                     with the conceptual translation"
                     /codon_start=1
                     /db_xref="PID:e25011"
                     /db_xref="PID:g1334904"
                     /db_xref="SWISS-PROT:P30119"
                     /translation="NERLPFSWPTCAKRQPPSTAVMLKCKQPGARFIHGAVHLPSGQI
                     VFHTIHSPTLASALGLPGENVPIPALFRASGLNVRESLPMTNMRAPIISLARLILAPN
                     PYILEGQLTVGMTQDNGIPVLFARPVIEVKSGPESNIKASSQLMIAEDSCLNQIAPFS
                     ASEHPAFSMVESVKRVRVDEGANTRRTIRDILEIPVTVLSSLQLSPTKSILKKAPEPP
                     PPEPQATFDATPYARIFYDIGRQVPKLGNAPAAQVSNVLIANRSHNSLRLVPNPDLLP
                     LQHLYLKHVVLKSLNLENIVQDFEAIFTSPSDTISEAETKAFEKLVEQAKNTVENIVF
                     CLNSICSTSTLPDVVPDVNNPNISLALEKYFLMFPPSGTIMRNVRFATPIVRLLCQGA
                     ELGTMAQFLGKYIKVKKETGMYTLVKLYYLLRI"
misc_feature         complement(140902)
                     /note="polyA signal: AATAAA, 3' end of 2.5kb late RNA"
CDS                  complement(140916. .143036)
                     /note="BXLF2 late reading frame, encodes gp85; homologous
                     to RF 37 VZV and glycoprotein H of HSV (gpIII of VZV)"
                     /codon_start=1
                     /db_xref="PID:g1334905"
                     /db_xref="SWISS-PROT:P03231"
                     /translation="MQLLCVFCLVLLWEVGAASLSEVKLHLDIEGHASHYTIPWTELM
                     AKVPGLSPEALWREANVTEDLASMLNRYKLIYKTSGTLGIALAEPVDIPAVSEGSMQV
                     DASKVHPGVISGLNSPACMLSAPLEKQLFYYIGTMLPNTRPHSYVFYQLRCHLSYVAL
                     SINGDKFQYTGAMTSKFLMGTYKRVTEKGDEHVLSLVFGKTKDLPDLRGPFSYPSLTS
                     AQSGDYSLVIVTTFVHYANFHNYFVPNLKDMFSRAVTMTAASYARYVLQKLVLLEMKG
                     GCREPELDTETLTTMFEVSVAFFKVGHAVGETGNGCVDLRWLAKSFFELTVLKDIIGI
                     CYGATVKGMQSYGLERLAAMLMATVKMEELGHLTTEKQEYALRLATVGYPKAGVYSGL
                     IGGATSVLLSAYNRHPLFQPLHTVMRETLFIGSHVVLRELRLNVTTQGPNLALYQLLS
                     TALCSALEIGEVLRGLALGTESGLFSPCYLSLRFDLTRDKLLSMAPQEATLDQAAVSN
                     AVDGFLGRLSLEREDRDAWHLPAYKCVDRLDKVLMIIPLINVTFIISSDREVRGSALY
                     EASTTYLSSSLFLSPVIMNKCSQGAVAGEPRQIPKIQNFTRTQKSCIFCGFALLSYDE
                     KEGLETTTYITSQEVQNSILSSNYFDFDNLHVHYLLLTTNGTVMEIAGLYEERAHVVL
                     AIILYFIAFALGIFLVHKIVMFFL"
misc_feature         140970
                     /note="polyA signal: AATAAA"
misc_feature         complement(141286)
                     /note="polyA signal: AATAAA"
promoter             142589
                     /note="TATA: GATAAAA"
misc_feature         142740
                     /note="BAM: Bam H1 T/X"
CDS                  complement(143038. .144861)
                     /note="BXLF1 early reading frame, thymidine kinase
                     (Littler et al, 1986). Weak homology to RF 36 VZV and HSV
                     thymidine kinase. 4.0kb early RNA presumably encodes the
                     TK. Also a 2.2kb late RNA here."
                     /codon_start=1
                     /db_xref="PID:g1334907"
                     /db_xref="SWISS-PROT:P03177"
                     /translation="MAGFPGKEAGPPGGWRKCQEDESPENERHENFYAEIDDFAPSVL
                     TPTGSDSGAGEEDDDGLYQVPTHWPPLMAPTGLSGERVPCRTQAAVTSNTGNSPGSRH
                     TSCPFTLPRGAQPPAPAHQKPTAPTPKPRSRECGPSKTPDPFSWFRKTSCTEGGADST
                     SRSFMYQKGFEEGLAGLGLDDKSDCESEDESNFRRPSSHSALKQKNGGKGKPSGLFEH
                     LAAHGREFSKLSKHAAQLKRLSGSVMNVLNLDDAQDTRQAKAQRKESMRVPIVTHLTN
                     HVPVIKPACSLFLEGAPGVGKTTMLNHLKAVFGDLTIVVPEPMRYWTHVYENAIKAMH
                     KNVTRARHGREDTSAEVLACQMKFTTPFRVLASRKRSLLVTESGARSVAPLDCWILHD
                     RHLLSASVVFPLMLLRSQLLSYSDFIQVLATFTADPGDTIVWMKLNVEENMRRLKKRG
                     RKHESGLDAGYLKSVNDAYHAVYCAWLLTQYFAPEDIVKVCAGLTTITTVCHQSHTPI
                     IRSGVAEKLYKNSIFSVLKEVIQPFRADAVLLEVCLAFTRTLAYLQFVLVDLSEFQDD
                     LPGCWTEIYMQALKNPAIRSQFFDWAGLSKVISDFERGNRD"
promoter             complement(143310)
                     /note="TATA: TATAAGA ECL2 late promoter before BXLF2,
                     gives 2.5kb late RNA"
misc_feature         143608
                     /note="polyA signal: AATAAA"
misc_feature         144791
                     /note="ACCEPT: TCTTTCGTTTTCAGG poss. acceptor before
                     BXRF1"
CDS                  144860. .145606
                     /note="BXRF1 late reading frame, homologous to RF 35 VZV.
                     Basic (core?) protein."
                     /codon_start=1
                     /db_xref="PID:g1334906"
                     /db_xref="SWISS-PROT:P03232"
                     /translation="MDPTRGLCALSTHDLAKFHSLPPARKAAGKRAHLRCYSKLLSLK
                     SWEQLASFLSLPPGPTFTDFRLFFEVTLGRRIADCVVVALQPYPRCYIVEFKTAMSNT
                     ANPQSVTRKAQRLEGTAQLCDCANFLRTSCPPVLGSQGLEVLAALVFKNQRSLRTLQV
                     EFPALGQKTLPTSTTGLLNLLSRWQDGALRARLDRPRPTAQGHRPRTHVGPKPSQLTA
                     RVPRSARAGRAGGRKGQVGAVGQVCPGAQK"
misc_feature         144862
                     /note="BAM: Bam H1 X/V"
misc_feature         144945
                     /note="DONOR: CAGGTAAGC possible donor at 3' BXRF1"
promoter             complement(145135)
                     /note="TATA: TATAACA before BXLF1"
promoter             145302
                     /note="TATA: TATTTAA before BVRF1, potential promoter for
                     1.9kb early RNA"
CDS                  145416. .147128
                     /note="BVRF1 early reading frame, homologous to RF 34 VZV"
                     /codon_start=1
                     /db_xref="PID:g1334908"
                     /db_xref="SWISS-PROT:P03233"
                     /translation="MALSGHVLIDPARLPRDTGPELMWAPSLRNSLRVSPEALELAER
                     EAERARSERWDRCAQVLKNRLLRVELDGIMRDHLARAEEIRQDLDAVVAFSDGLESMQ
                     VRSPSTGGRSAPAPPSPSPAQPFTRLTGNAQYAVSISPTDPPLMVAGSLAQTLLGNLY
```

GNINQWVPSFGPWYRTMSANAMQRRVFPKQLRGNLNFTNSVSLKLMTEVVAVLEGTTQ
DFFSDVRHLPDLQAALILSVAYLLLQGGSSHQQRPLPASREELLELGPESLEKIIADL
KAKSPGGNFMILTSGNKEARQSIAPLNRQAAYPPGTFADNKIYNLFVGAGLLPTTAAL
NVPGAAGRDRDLVYRIANQIFGEDVPPFSSHQWNLRVGLAALEALMLVYTLCETANLA
EAATRRLHLSSLLPQAMQRRKPAMASAGMPGAYPVQTLFRHGELFRFIWAHYVRPTVA
ADPQASISSLFPGLVLLALELKLMDGQAPSHYAINLTGQKFDTLFEIINQKLLFHDPA
AMLAARTQLRLAFEDGVGVALGRPSPMLAAREILERQFSASDDYDRLYFLTLGYLASP
VAPS"

misc_feature complement(146926)
/note="polyA signal: AATAAA"

misc_feature 147167
/note="DONOR: AAGGTAAAT possible donor"

misc_feature 147170
/note="polyA signal: AATAAA, 3' end of 2.4kb late and 1.9kb early RNAs"

promoter 147721
/note="TATA: TATTTAT before BVRF2, potential promoter for 2.1kb early RNA"

CDS 147927. .149744
/note="BVRF2 early reading frame, N-terminus homologous to RF 33 VZV"
/codon_start=1
/db_xref="PID:g1334909"
/db_xref="SWISS-PROT:P03234"
/translation="MVQAPSVYVCGFVERPDAPPKDACLHLDPLTVKSQLPLKKPLPL
TVEHLPDAPVGSVFGLYQSRAGLFSAASITSGDFLSLLDSIYHDCDIAQSQRLPLPRE
PKVEALHAWLPSLSLASLHPDIPQTTADGGKLSFFDHVSICALGRRRGTTAVYGTDLA
WVLKHFSDLEPSIAAQIENDANAAKRESGCPEDHPLPLTKLIAKAIDAGFLRNRVETL
RQDRGVANIPAESYLKASDAPDLQKPDKALQSPPPASTDPATMLSGNAGEGATACGGS
AAAGQDLISVPRNTFMTLLQTNLDNKPPRQTPLPYAAPLPPFSHQAIATAPSYGPGAG
AVAPAGGYFTSPGGYYAGPAGGDPGAFLAMDAHTYHPHPHPPPAYFGLPGLFGPPPPV
PPYYGSHLRADYVPAPSRSNKRKRDPEEDEEGGGLFPGEDATLYRKDIAGLSKSVNEL
QHTLQALRRETLSYGHTGVGYCPQQGPCYTHSGPYGFQPHQSYEVPRYVPHPPPPPTS
HQAAQAQPPPPGTQAPEAHCVAESTIPEAGAAGNSGPREDTNPQQPTTEGHHRGKKLV
QASASGVAQSKEPTTPKAKSVSAHLKSIFCEELLNKRVA"

misc_feature 148007
/note="BAM: Bam H1 V/d"

promoter 148620
/note="TATA: TATTTAA ECR1 late promoter before BdRF1, gives 1.2kb late RNA"

CDS 148707. .149744
/note="BdRF1 reading frame; this is the C terminus of BVRF2"
/codon_start=1
/db_xref="PID:g1334910"
/db_xref="SPTREMBL:Q66541"
/translation="MLSGNAGEGATACGGSAAAGQDLISVPRNTFMTLLQTNLDNKPP
RQTPLPYAAPLPPFSHQAIATAPSYGPGAGAVAPAGGYFTSPGGYYAGPAGGDPGAFL
AMDAHTYHPHPHPPPAYFGLPGLFGPPPPVPPYYGSHLRADYVPAPSRSNKRKRDPEE
DEEGGGLFPGEDATLYRKDIAGLSKSVNELQHTLQALRRETLSYGHTGVGYCPQQGPC
YTHSGPYGFQPHQSYEVPRYVPHPPPPPTSHQAAQAQPPPPGTQAPEAHCVAESTIPE
AGAAGNSGPREDTNPQQPTTEGHHRGKKLVQASASGVAQSKEPTTPKAKSVSAHLKSI
FCEELLNKRVA"

misc_feature 149115
/note="BAM: Bam H1 d/I"

misc_feature 149727
/note="polyA signal: AATAAA, 3' end of 2.1kb early and 1.2kb late RNAs"

misc_feature complement(149758)
/note="polyA signal: AATAAA, 3' end of 1.0kb late, 1.5kb late and 1.8kb late RNAs"

CDS complement(149779. .150525)
/note="BILF2 late reading frame 11xNXT/S"
/codon_start=1
/db_xref="PID:g1334911"
/db_xref="SWISS-PROT:P03218"
/translation="MTHLVLLLCCCVGSVCAFFSDLVKFENVTAHAGARVNLTCSVPS
NESVSRIELGRGYTPGDGQLPLAVATSNNGTHITNGGYNYSLTLEWVNDSNTSVSLII
PNVTLAHAGYYTCNVTLRNCSVASGVHCNYSAGEEDDQYHANRTLTQRMHLTVIPATT
IAPTTLVSHTTSTSHRPHRRPVSKRPTHKPVTLGPFPIDPWRPKTTWVHWALLLITCA
VVAPVLLIIIISCLGWLAGWGRRRKGWIPL"

promoter complement(150571)
/note="TATA: TATTTAG before BILF2. Potential promoter for 1.0kb late RNA."

repeat_region 151236. .151618
/note="repetitive sequence 3X"

misc_feature 151767
/note="polyA signal: AATAAA"

promoter complement(151780)
/note="TATA: CATAAAA"

misc_feature 152012. .152013
/note="DEL: B95-8 deletion with respect to Raji"

CDS complement(152161. .153099)
/note="BILF1 reading frame, membrane protein, 3xNXS /T"
/codon_start=1
/db_xref="PID:g1334912"
/db_xref="SWISS-PROT:P03208"
/translation="MLSTMAPGSTVGTLVANMTSVNATEDACTKSYSAFLSGMTSLLL
VLLILLTLAGILFIIFVRKLVHRMDVWLIALLIELLLWVLGKMIQEFSSTGLCLLTQN
MMFLGLMCSVWTHLGMALEKTLALFSRTPKRTSHRNVCLYLMGVFCLVLLLIIILLIT
MGPDANLNRGPNMCREGPTKGMHTAVQGLKAGCYLLAAVLIVLLTVIIIWKLLRTKFG
RKPRLICNVTFTGLICAFSWFMLSLPLLFLGEAGSLGFDCTESLVARYYPGPAACLAL
LLIILYAWSFSHFMDSLKNQVTVTARYFRRVPSQST"

promoter 152230
/note="TATA: CATAAAA"

misc_feature 153259
/note="polyA signal: AATAAA"

misc_feature 153637
/note="HPN: 22bp 2-fold symmetric"

misc_feature complement(153690)
/note="DONOR: AAAGTGAGG possible donor"

CDS complement(153699. .156746)
/note="BALF5 DNA polymerase (early), homologous to many DNA polymerases, CMV HFLF2 and RF 28 VZV. 4.5kb early RNA apparently encodes BALF5, RNA ends unknown."
/codon_start=1
/db_xref="PID:g1334913"
/db_xref="SWISS-PROT:P03198"
/translation="MSGGLFYNPFLRPNKGLLKKPDKEYLRLIPKCFQTPGAAGVVDV
RGPQPPLCFYQDSLTVVGGDEDGKGMWWRQRAQEGTARPEADTHGSPLDFHVYDILET
VYTHEKCAVIPSDKQGYVVPCGIVIKLLGRRKADGASVCVNVFGQQAYFYASAPQGLD
VEFAVLSALKASTFDRRTPCRVSVEKVTRRSIMGYGNHAGDYHKITLSHPNSVCHVAT
WLQDKHGCRIFEANVDATRRFVLDNDFVTFGWYSCRRAIPRLQHRDSYAELEYDCEVG
DLSVRREDSSWPSYQALAFDIECLGEEGFPTATNEADLILQISCVLWSTGEEAGRYRR
ILLTLGTCEDIEGVEVYEFPSELDMLYAFFQLIRDLSVEIVTGYNVANFDWPYILDRA
RHIYSINPASLGKIRAGGVCEVRRPHDAGKGFLRANTKVRITGLIPIDMYAVCRDKLS
LSDYKLDTVARHLLGAKKEDVHYKEIPRLFAAGPEGRRRLGMYCVQDSALVMDLLNHF
VIHVEVAEIAKIAHIPCRRVLDDGQQIRVFSCLLAAAQKENFILPMPSASDRDGYQGA
TVIQPLSGFYNSPVLVVDFASLYPSIIQAHNLCYSTMITPGEEHRLAGLRPGEDYESF
RLTGGVYHFVKKHVHESFLASLLTSWLAKRKAIKKLLAACEDPRQRTILDKQQLAIKC
TCNAVYGFTGVANGLFPCLSIAETVTLQGRTMLERAKAFVEALSPANLQALAPSPDAW
APLNPEGQLRVIYGDTDSLFIECRGFSESETLRFADALAAHTTRSLFVAPISLEAEKT
FSCLMLITKKRYVGVLTDGKTLMKGVELVRKTACKFVQTRCRRVLDLVLADARVKEAA
SLLSHRPFQESFTQGLPVGFLPVIDILNQAYTDLREGRVPMGELCFSTELSRKLSAYK
STQMPHLAVYQKFVERNEELPQIHDRIQYVFVEPKGGVKGARKTEMAEDPAYAERHGV
PVAVDHYFDKLLQGAANILQCLFDNNSGAALSVLQNFTARPPF"

misc_feature 154747
/note="BAM: Bam H1 I/A"

misc_feature complement(156707)
/note="polyA signal: AATAAA; 3' end of 2.5kb late (gB) RNA and 1.8kb late RNA"

CDS complement(156749. .159322)
/note="BALF4 late reading frame 9xNXT/S homologous to HSV1 glycoprotein B (Pellet et al, 1985), CMV HFLF1 and RF 31 VZV (gpII)"
/codon_start=1
/db_xref="PID:g1334914"
/db_xref="SWISS-PROT:P03188"
/translation="MTRRRVLSVVVLLAALACRLGAQTPEQPAPPATTVQPTATRQQT
SFPFRVCELSSHGDLFRFSSDIQCPSFGTRENHTEGLLMVFKDNIIPYSFKVRSYTKI
VTNILIYNGWYADSVTNRHEEKFSVDSYETDQMDTIYQCYNAVKMTKDGLTRVYVDRD
GVNITVNLKPTGGLANGVRRYASQTELYDAPGWLIWTYRTRTTVNCLITDMMAKSNSP
FDFFVTTTGQTVEMSPFYDGKNKETFHERADSFHVRTNYKIVDYDNRGTNPQGERRAF
LDKGTYTLSWKLENRTAYCPLQHWQTFDSTIATETGKSIHFVTDEGTSSFVTNTTVGI
ELPDAFKCIEEQVNKTMHEKYEAVQDRYTKGQEAITYFITSGGLLLAWLPLTPRSLAT
VKNLTELTTPTSSPPSSPSPPAPSAARGSTPAAVLRRRRRDAGNATTPVPPTAPGKSL

GTLNNPATVQIQFAYDSLRRQINRMLGDLARAWCLEQKRQNMVLRELTKINPTTVMSS
IYGKAVAAKRLGDVISVSQCVPVNQATVTLRKSMRVPGSETMCYSRPLVSFSFINDTK
TYEGQLGTDNEIFLTKKMTEVCQATSQYYFQSGNEIHVYNDYHHFKTIELDGIATLQT
FISLNTSLIENIDFASLELYSRDEQRASNVFDLEGIFREYNFQAQNIAGLRKDLDNAV
SNGRNQFVDGLGELMDSLGSVGQSITNLVSTVGGLFSSLVSGFISFFKNPFGGMLILV
LVAGVVILVISLTRRTRQMSQQPVQMLYPGIDELAQQHASGEGPGINPISKTELQAIM
LALHEQNQEQKRAAQRAAGPSVASRALQAARDRFPGLRRRRYHDPETAAALLGEAETE
F"

promoter 158204
/note="TATA: TATAAAA"

CDS complement(159312. .161678)
/note="BALF3 reading frame; Protein sequence is in conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e24996"
/db_xref="PID:g1334915"
/db_xref="SWISS-PROT:P25939"
/translation="GRRGVLIGPLLRPGGQRPRNPGDHCLQRDRVDGGGRSGLSGGSV
GWARGSPGQLYQRGRAGQHRVAGLGRAAAGWRRAGLRGQTQAPSGHRSPRTRGLETPG
AAMSGLLAAAYSQVYALAVELSVCTRLDPRSLDVAAVVRNAGLLAELEAILLPRLRRQ
NDRACSALSLELVHLLENSREASAALLAPGRKGTRVPPLRTPSVAYSVEFYGGHKVDV
SLCLINDIEILMKRINSVFYCMSHTMGLESLERALDLLGRFRGVSPIPDPRLYITSVP
CWRCVGELMVLPNHGNPSTAEGTHVSCNHLAVPVNPEPVSGLFENEVRQAGLGHLLEA
EEKARPGGPEEGAVPGPGRPEAEGATRALDTYNVFSTVPPEVAELSELLYWNSGGHAI
GATGQGEGGGHSRLSALFARERRLALVRGACEEALAGARLTHLFDAVAPGATERLFCG
GVYSSSGDAVEALKADCAAAFTAHPQYRAILQKRNELYTRLNRAMQRLGRGEEEASRE
SPEVPRPAGAREPGPSGALSDALKRKEQYLRQVATEGLAKLQSCLAQQSETLTETLCL
RVWGDVVYWELARMRNHFLYRRAFVSGPWEDRRAGEGAAFENSKYIKTHLFTQTLSSE
HLHALTHSLYTFITGPLAEESGLFPPPSNVALARCCDAAGTLPHQKAFLTSLIWPGIE
PSDWIETSFNSFYSVPGGSLASSQQILCRALREAVLTVSLYNKTWGRSLILRRADAVS
PGQALPPDGLYLTYDSDRPLILLYKGRGWVFKDLYALLYLHLQMRDDSA"

promoter complement(159370)
/note="TATA: TATTTAA ECL1 late promoter before BALF4, gives 2.5kb late RNA"

misc_feature 160966
/note="polyA signal: AATAAA"

misc_feature 160990
/note="polyA signal: ATTAAA"

misc_feature complement(161013)
/note="polyA signal: AATAAA, presumed end of 3.9kb early RNA"

CDS complement(161384. .164770)
/note="BALF2 early reading frame, homologous to RF 29 VZV and major DNA binding protein HSV. 3.9kb early RNA"
/codon_start=1
/db_xref="PID:g1334916"
/db_xref="SWISS-PROT:P03227"
/translation="MQGAQTSEDNLGSQSQPGPCGYIYFYPLATYPLREVATLGTGYA
GHRCLTVPLLCGITVEPGFSINVKALHRRPDPNCGLLRATSYHRDIYVFHNAHMVPPI
FEGPGLEALCGETREVFGYDAYSALPRESSKPGDFFPEGLDPSAYLGAVAITEAFKER
LYSGNLVAIPSLKQEVAVGQSASVRVPLYDKEVFPEGVPQLRQFYNSDLSRCMHEALY
TGLAQALRVRRVGKLVELLEKQSLQDQAKVAKVAPLKEFPASTISHPDSGALMIVDSA
ACELAVSYAPAMLEASHETPASLNYDSWPLFADCEGPEARVAALHRYNASLAPHVSTQ
IFATNSVLYVSGVSKSTGQGKESLFNSFYMTHGLGTLQEGTWDPCRRPCFSGWGGPDV
TGTNGPGNYAVEHLVYAASFSPNLLARYAYYLQFCQGQKSSLTPVPETGSYVAGAAAS
PMCSLCEGRAPAVCLNTLFFRLRDRFPPVMSTQRRDPYVISGASGSYNETDFLGNFLN
FIDKEDDGQRPDDEPRYTYWQLNQNLLERLSRLGIDAEGKLEKEPHGPRDFVKMFKDV
DAAVDAEVVQFMNSMAKNNITYKDLVKSCYHVMQYSCNPFAQPACPIFTQLFYRSLLT
ILQDISLPICMCYENDNPGLGQSPPEWLKGHYQTLCTNFRSLAIDKGVLTAKEAKVVH
GEPTCDLPDLDAALQGRVYGRRLPVRMSKVLMLCPRNIKIKNRVVFTGENAALQNSFI
KSTTRRENYIINGPYMKFLNTYHKTLFPDTKLSSLYLWHNFSRRRSVPVPSGASAEEY
SDLALFVDGGSRAHEESNVIDVVPGNLVTYAKQRLNNAILKACGQTQFYISLIQGLVP
RTQSVPARDYPHVLGTRAVESAAAYAEATSSLTATTVVCAATDCLSQVCKARPVVTLP
VTINKYTGVNGNNQIFQAGNLGYFMGRGVDRNLLQAPGAGLRKQAGGSSMRKKFVFAT
PTLGLTVKRRTQAATTYEIENIRAGLEAIISQKQEEDCVFDVVCNLVDAMGEACASLT
RDDAEYLLGRFSVLADSVLETLATIASSGIEWTAEAARDFLEGVWGGPGAAQDNFISV
AEPVSTASQASAGLLLGGGGQGSGGRRKRRLATVLPGLEV"

misc_feature 163978. .166635
/note="DEL: deletion in Raji"

promoter complement(164814)
/note="TATA: CATTTAA before BALF2, presumed promoter for 3.9kb early RNA"

misc_feature complement(164851)
/note="polyA signal: AATAAA"

CDS complement(164855. .165517)
/note="BALF1 early reading frame, 0.7kb early RNA"
/codon_start=1
/db_xref="PID:g1334918"
/db_xref="SWISS-PROT:P03229"
/translation="MNLAIALDSPHPGLASYTILPRPFYHISLKPVSWPDETMRPAKS
TDSVFVRTPVEAWVAPSPPDDKVAESSYLMFRAMYAVFTRDEKDLPLPALVLCRLIKA
SLRKDRKLYAELACRTADIGGKDTHVRLIISVLRAVYNDHYDYWSRLRVVLCYTVVFA
VRNYLDDHKSAAFVLGAIAHYLALYRRLWFARLGGMPRSLRRQFPVTWALASLTDFLK
SL"

promoter 165442
/note="TATA: GATAAAA"

promoter 165466
/note="TATA: TATAAGA EDR1 early promoter before BARF1, gives 0.8kb early RNA"

CDS 165504. .166169
/note="BARF1 reading frame"
/codon_start=1
/db_xref="PID:g1334917"
/db_xref="SWISS-PROT:P03228"
/translation="MARFIAQLLLLASCVAAGQAVTAFLGERVTLTSYWRRVSLGPEI
EVSWFKLGPGEEQVLIGRMHHDVIFIEWPFRGFFDIHRSANTFFLVVTAANISHDGNY
LCRMKLGETEVTKQEHLSVVKPLTLSVHSERSQFPDFSVLTVTCTVNAFPHPHVQWLM
PEGVEPAPTAANGGVMKEKDGSLSVAVDLSLPKPWHLPVTCVGKNDKEEAHGVYVSGY
LSQ"

promoter complement(165713)
/note="TATA: TATAAAG before BALF1"

misc_feature 166165
/note="polyA signal: AATAAA 3' end of 0.8kb early RNA. Also 1kb late RNA in this region."

promoter 166469. .166475
/note="TATA: TTATTTT"

mRNA 166498. .166916
/note="Exon 1 of terminal protein RNA"

misc_feature 166561. .166563
/note="Likely initiator met of terminal protein"

misc_feature 166614
/note="BAM: Bam H1 A/Nhet"

misc_feature complement(166946)
/note="polyA signal: AATAAA These 2 polyA sites are 3' end of 0.8kb early, 2.5kb late and 2.5kb latent RNAs"

misc_feature complement(166950)
/note="polyA signal: AATAAA"

CDS complement(166998. .167303)
/note="BNLF2b reading frame; Protein sequence is in conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25001"
/db_xref="PID:g1334919"
/db_xref="SPTREMBL:Q04362"
/translation="EATMRPGRPLAGFYATLRRSFRRMSKRSKNKAKKERVPVEDRPP
TPMPTSQRLIRRNALGGGVRPDAEDCIQRFHPLEPALGVSTKNFDLLSLRCELGWCG"

CDS complement(167304. .167486)
/note="BNLF2a reading frame"
/codon_start=1
/db_xref="PID:g1334920"
/db_xref="SPTREMBL:Q04361"
/translation="MVHVLERALLEQQSSACGLPGSSTETRPSHPCPEDPDVSRLRLL
LVVLCVLFGLLCLLLI"

promoter 167320
/note="TATA: CATAAAA"

promoter complement(167525)
/note="TATA: TATAAAA EDL2 early promoter before BNLF2a,b. Gives 0.8kb RNA"

mRNA complement(168163. .168965)
/note="BNLF1 coding part of exon c of latent membrane protein"

repeat_region 168399. .168574
/note="5 x 33bp repeats"

intron complement(168966. .169041)
/note="intervening sequence in BNLF1"

misc_feature complement(168966)

```
                 /note="ACCEPT: TTTTTTCAACGCAGT possible acceptor"
     misc_feature    complement(169041)
                     /note="DONOR: TTGGTAAGA donor"
     mRNA            complement(169042. .169128)
                     /note="BNLF1 exon b of latent membrane protein mRNA"
     intron          complement(169129. .169206)
                     /note="intervening sequence in BNLF1"
     misc_feature    complement(169129)
                     /note="ACCEPT: TCCTCTTTCCCCAGT acceptor"
     promoter        complement(169201)
                     /note="TATA: TATTACA EDL1A late promoter, gives 2.5kb late
                     RNA"
     misc_feature    complement(169206)
                     /note="DONOR: TGAGTAAGT donor"
     mRNA            complement(169207. .169474)
                     /note="BNLF1 exon a of latent membrane protein mRNA"
     promoter        complement(169546)
                     /note="TATA: TACATAAGC EDL1 promoter before BNLF1 gives
                     2.5kb latent RNA (LMP)"
     repeat_region   170094. .170631
                     /note="terminal repeat 1 538bp"
     repeat_region   170632. .171154
                     /note="terminal repeat 2 523bp"
     repeat_region   171155. .171692
                     /note="terminal repeat 3 538bp"
     repeat_region   171693. .172231
                     /note="terminal repeat 4 538bp"
BASE COUNT    34054 a  52511 c  50755 g  34961 t
ORIGIN


  Query Match             6.2%;  Score 151.6;  DB 17;  Length 172281;
  Best Local Similarity   53.2%;  Pred. No. 1.7e-19;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || |||||||  | ||| || |  |  |||||||  | ||| |  |  |  |||||||
Db 108853 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 108794

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || ||  || |    | ||| ||| |  | ||  || |  |||  | || |  |
Db 108793 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 108734

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          |  |  |  | ||   |||| |||| |||||  | ||  | ||  |  | || ||| ||
Db 108733 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 108674

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || ||||||||    | ||  || ||    |    | ||| ||||  ||||  || |
Db 108673 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 108614

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || ||||    ||||       |     ||| | |  | || ||| | |||
Db 108613 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 108554

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  |||||||  | ||  | ||| || || ||| |||
Db 108553 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 108494

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |||  || |||   |  | || ||| ||||| |  || || | ||  | |||||||
Db 108493 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 108434

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |     |  |     |    ||  || || || |  || || || |  | |||||  ||
Db 108433 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 108374

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
            | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db 108373 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 108314

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
           ||| ||     || |  |||||||     || |  |   | | ||  || |  |  |

Db 108313 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 108254

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
               || |     |   ||| |||   |    |  ||  |    || || ||
Db 108253 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 108202


RESULT  12
HS4B958RAJ/c
LOCUS       HS4B958RAJ 184113 bp    DNA             VRL       12-APR-1996
DEFINITION  Epstein-Barr virus, artifactual joining of B95-8 complete genome
            and the sequences from Raji of the large deletion found in B95-8.
ACCESSION   M80517 M75989
NID         g330330
KEYWORDS    .
SOURCE      Human herpesvirus 4 DNA.
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (sites)
  AUTHORS   Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J.,
            Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C.,
            Tuffnell,P.S. and Barrell,B.G.
  TITLE     DNA sequence and expression of the B95-8 Epstein-Barr virus genome
  JOURNAL   Nature 310 (5974), 207-211 (1984)
  MEDLINE   84270667
REFERENCE   2  (sites)
  AUTHORS   Parker,B.D., Bankier,A., Satchwell,S., Barrell,B. and Farrell,P.J.
  TITLE     Sequence and transcription of Raji Epstein-Barr virus DNA spanning
            the B95-8 deletion region
  JOURNAL   Virology 179 (1), 339-346 (1990)
  MEDLINE   91021036
REFERENCE   3  (sites)
  AUTHORS   Sample,J., Brooks,L., Sample,C., Young,L., Rowe,M., Gregory,C.,
            Rickinson,A. and Kieff,E.
  TITLE     Restricted Epstein-Barr virus protein expression in Burkitt
            lymphoma is due to a different Epstein-Barr nuclear antigen 1
            transcriptional initiation site
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88 (14), 6343-6347 (1991)
  MEDLINE   91296817
REFERENCE   4  (bases 1 to 184113)
  AUTHORS   Jenson,H.B.
  TITLE     GenBank Curator Program
  JOURNAL   Unpublished (1992)
COMMENT     The B95-8 genome (V01555) has a large deletion in the right side of
            the genome which has been sequenced in Raji (M35547).  These
            sequences have been joined to form an extended and more complete,
            although artifactual, EBV sequence.
            For features, refer to feature tables of V01555 and M35547.
FEATURES             Location/Qualifiers
     source          1. .184113
                     /organism="Human herpesvirus 4"
                     /db_xref="taxon:10376"
     misc_feature    1. .152008
                     /note="B95-8 sequences (corresponds to 1-152,008 of
                     V01555)"
     misc_feature    152009. .152012
                     /note="Overlap of B95-8 and Raji sequences at B95-8
                     deletion point (corresponds to 152,009-152,012 in V01555,
                     and 1-4 in M35547)"
     misc_feature    153013. .163839
                     /note="Raji sequences (corresponds to 5-11,831 of M35547)"
     misc_feature    163840. .163843
                     /note="Overlap of B95-8 and Raji sequences at B95-8
                     deletion point (corresponds to 152,009-152,012 of V01555,
                     and 11,832-11,835 of M35547)"
     misc_feature    163844. .184113
                     /note="B95-8 sequences (corresponds to 152,013-172,282 of
                     V01555)"
BASE COUNT    36002 a  55824 c  54622 g  37665 t
ORIGIN
```

```
  Query Match             6.2%;  Score 151.6;  DB 17;  Length 184113;
  Best Local Similarity   53.2%;  Pred. No. 1.7e-19;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || ||||||| | ||| || | | ||||||| | ||| | | | |||||||
Db 108853 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 108794

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || || || | | ||| ||| | | || || | | |||| | || | |
Db 108793 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 108734

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          | | | | || |||| |||| ||||| | || | || | | || ||| ||
Db 108733 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 108674

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || |||||||| | || || || | ||| |||| |||| || |
Db 108673 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 108614

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || |||| |||| | ||| | | | || ||| | |||
Db 108613 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 108554

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  ||||||| | ||  | ||| || || ||| |||
Db 108553 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 108494

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |||  || |||   | | || ||| ||||| |  || || | ||  | |||||||
Db 108493 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 108434

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |     | |     |    ||  || || || |  || || || | |  | ||||| ||
Db 108433 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 108374

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
            | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db 108373 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 108314

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
          |||  ||     ||  |  ||||||||     || |  |   | | ||  || |  |  |
Db 108313 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 108254

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
            || |     |  ||| |||   |     |  ||  |    || || ||
Db 108253 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 108202


RESULT  13
HS4ULIR3/c
LOCUS       HS4ULIR3     1150 bp    DNA             VRL       15-JUN-1990
DEFINITION  epstein-barr virus simple repeat array (ir3).
ACCESSION   J02079
NID         g330445
KEYWORDS    repeat region.
SOURCE      ebv (epstein barr virus) from human.
  ORGANISM  Human herpesvirus 4
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE   1  (bases 1 to 1150)
  AUTHORS   Heller,M., van Santen,V.L. and Keiff,E.
  TITLE     simple repeat sequence in epstein-barr virus dna is transcribed in
            latent and productive infections
  JOURNAL   J. Virol. 44, 311-320 (1982)
  MEDLINE   83059881
FEATURES             Location/Qualifiers
     source          1. .1150
                     /organism="Human herpesvirus 4"
                     /db_xref="taxon:10376"
BASE COUNT      302 a    169 c    633 g     46 t
ORIGIN
```

```
  Query Match             6.2%;  Score 151.6;  DB 17;  Length 1150;
  Best Local Similarity   53.2%;  Pred. No. 2.8e-19;
  Matches  347;  Conservative    0;  Mismatches  299;  Indels    6;  Gaps    1;

Qy   1500 CTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGCTGCCCCAA 1559
          || ||||||| | ||| || | | ||||||| | ||| | | | |||||||
Db    908 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTC 849

Qy   1560 CTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGCCCCAGTCC 1619
          || | || || || | | ||| ||| | | || || | | |||| | || | |
Db    848 CTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTC 789

Qy   1620 CACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGCTCCTACTA 1679
          | | | | || |||| |||| ||||| | || | || | | || ||| ||
Db    788 CTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTG 729

Qy   1680 CTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCACCGCAG 1739
          || |||||||| | || || || | ||| |||| |||| || |
Db    728 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 669

Qy   1740 CCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTACCGCT---- 1796
          | ||| || |||| |||| | ||| | | | || ||| | |||
Db    668 CTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCTCCTG 609

Qy   1796 --GCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCCAG 1853
            | ||  | ||| |  || |  ||||||| | ||  | ||| || || ||| |||
Db    608 CCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 549

Qy   1854 AAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGCCC 1913
            |||  || |||   | | || ||| ||||| |  || || | ||  | |||||||
Db    548 CTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCC 489

Qy   1914 CCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCCCA 1973
          |     | |     |    ||  || || || |  || || || | |  | ||||| ||
Db    488 CTCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTG 429

Qy   1974 ACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGCAC 2033
            | |  | | | ||    | ||  || |||| ||  |  |  ||||  || | | | | |
Db    428 CCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCCCTC 369

Qy   2034 ATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACCTG 2093
          |||  ||     ||  |  ||||||||     || |  |   | | ||  || |  |  |
Db    368 CTGCCCCTCCTGCCCCTCCTGCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCTGCCC 309

Qy   2094 AAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCAACCCTTGCGT 2145
            || |     |  ||| |||   |     |  ||  |    || || ||
Db    308 CTCCTGCTCCTGCCCCTCCCGCTCCTGCTCCTGCTCCTGTTCCACCGTGGGT 257


RESULT  14
D88733
LOCUS       D88733       3170 bp    DNA             VRL       19-MAY-1997
DEFINITION  Equine herpesvirus 1 DNA for membrane glycoprotein, complete cds.
ACCESSION   D88733
NID         g2114320
KEYWORDS    membrane glycoprotein.
SOURCE      Equine herpesvirus 1 (strain:HH1) DNA.
  ORGANISM  Equine herpesvirus 1
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 3170)
  AUTHORS   Kirisawa,R., Kobayashi,T., Kawakami,Y. and Iwai,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-1996) to the DDBJ/EMBL/GenBank databases. Rikio
            Kirisawa, Rakuno Gakuen University, Faculty of Veterinary Medicine,
            Department of Veterinary Microbiology; Bunkyodai-Midorimachi 582,
            Ebetsu, Hokkaido 069, Japan (E-mail:r-kirisa@ca2.so-net.or.jp,
            Tel:+81-11-386-1112, Fax:+81-11-387-5890)
REFERENCE   2  (sites)
  AUTHORS   Kirisawa,R., Kobayashi,T., Kawakami,Y. and Iwai,H.
  TITLE     Nucleotide sequences of open reading frames 1, 24 and 71 of an
            attenuated equine herpesvirus-1
```

```
  JOURNAL   J. Equine Sci. 7, 79-87 (1996)
FEATURES             Location/Qualifiers
     source          1. .3170
                     /organism="Equine herpesvirus 1"
                     /strain="HH1"
                     /db_xref="taxon:10326"
     gene            118. .2718
                     /gene="ORF71"
     CDS             118. .2718
                     /gene="ORF71"
                     /codon_start=1
                     /product="membrane glycoprotein"
                     /db_xref="PID:d1020845"
                     /db_xref="PID:g2114321"
                     /translation="MGFIYARKLLLCMAVSIYAIGSTTTTETTTSSSSTSGSGQSTSS
                     GTTNSSSSPTTSPPTTSSSPPTSTHTSSPSSTSTQSSSTAATSSSAPSTASSTTSIPT
                     STSTETTTTTPTASTTTPTTTTAAPTTAATTTAVTTAASTAASTSAETTTATATATST
                     PTTTTPTSTTTTTATTTVPTTASTTTDTTTAATTTAATTTAATTTAATTTAATTTAAT
                     TTAATTTAATTTAATTTAATTTAATTSSATTAATTTAATTTAATTTAATTTAATTTAA
                     TTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTGSPTS
                     GSTSTTGASTSTPSASTATSATPTSTSTSAAATTSTPTPTSAATSAESTTEAPTSTPT
                     TDTTTPSEATTATTSPESTTVSASTTSATTTAFTTESHTSPDSSTGSTSTAEPSSTFT
                     LTPSTATPSTDQFTGSSASTESDSTDSSTVPTTGTESITESSSTTEASTNLGSSTYES
                     TEALETPDGNTTSGNTTPSPSPRTPSFADTQQTPDNGVSTQHTTINDHTTANAQKHAG
                     HHRGRAGGRRGSPQGGSHTTPHPDRLTPSPDDTYDDDTNHPNGRNNSIEIVPQLPPDR
                     PIIELGVATLRKNFMEASCTVETNSGLAIFWKIGNASVDAFNRGTTHTRLMRNGVPVY
                     ALVSTLRVPWLNVIPLTKITCAACPTNLVAGDGVDLNSCTTKSTTIPCPGQQRTHIFF
                     SAKGDRAVCITSELVSQPTITWSVGSDRLRNDGFSQTWYGIQPGVCGILRSEVRIHRT
                     TWRFGSTSKDYLCEVSASDSKTSDYKVLPNAHSTSNFALVAATTLTVTILCLLCCLYC
                     MLTRPRASVY"
     repeat_region   532. .555
     repeat_region   679. .828
     repeat_region   859. .1098
     polyA_signal    2799. .2804
BASE COUNT      847 a   1072 c    644 g    607 t
ORIGIN


  Query Match             5.9%;  Score 145.6;  DB 17;  Length 3170;
  Best Local Similarity   52.5%;  Pred. No. 3.3e-18;
  Matches  319;  Conservative    0;  Mismatches  289;  Indels    0;  Gaps    0;

Qy   1492 CACATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGC 1551
          |||| | ||  || |  | ||| |  || | || ||| || | || | | |  |||| ||
Db    603 CACAACTACGCCTACGTCCACAACAACTACTACAGCTACCACCACTGTTCCAACAACCGC 662

Qy   1552 TGCCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGC 1611
          | |  |||| | |   | |||| |  |   |  | || |  || |  | || |||||
Db    663 TTCTACAACAACTGATACGACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGC 722

Qy   1612 CCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGC 1671
            |   | || |  |||| | |   |  |  || || |  | ||  |  |||| |||||
Db    723 AACGACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGC 782

Qy   1672 TCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCC 1731
            | || || || ||  |     |  || | ||   ||||| | || | ||  | ||  |
Db    783 AACGACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAACGACCACCGCGGC 842

Qy   1732 CACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTAC 1791
           ||  |  || || | ||  | |   ||   | | |      ||  |||  | ||| ||||
Db    843 TACTACTTCCTCTGCAACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTAC 902

Qy   1792 CGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCC 1851
          | |  || |  |  |  || || | ||| || ||||||| ||  | || || || |   |
Db    903 CACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTAC 962

Qy   1852 AGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGC 1911
            | | |  ||||   |||  |  |  | || |  | || ||| | ||  || | |
Db    963 CACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTAC 1022

Qy   1912 CCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCC 1971
          | ||| | |  |   |||     |  | ||  |  || |  ||||| |  |  | ||  |
Db   1023 CACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTAC 1082

Qy   1972 CAACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGC 2031
          || |||| | |  ||   | |     || |  || |   | || |  |  |||| |  ||
Db   1083 CACCACCGCGGCTACCACAACGGGGTCTCCAACCTCTGGTTCAACATCTACTACAGGGGC 1142

Qy   2032 ACATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACC 2091
            |  ||| | || ||||| || | |  ||  |  |||   |  || | ||  || | |
Db   1143 TTCCACGTCCACCCCCTCAGCTTCCACTGCCACATCTGCCACTCCCACATCGACGTCAAC 1202

Qy   2092 TGAAGATG 2099
            || ||
Db   1203 ATCAGCTG 1210


RESULT  15
D88734
LOCUS       D88734       3173 bp    DNA             VRL       19-MAY-1997
DEFINITION  Equine herpesvirus 1 DNA for membrane glycoprotein, complete cds.
ACCESSION   D88734
NID         g2114322
KEYWORDS    membrane glycoprotein.
SOURCE      Equine herpesvirus 1 (strain:BK343, isolate:3F clone) DNA.
  ORGANISM  Equine herpesvirus 1
            Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Alphaherpesvirinae; Varicellovirus.
REFERENCE   1  (bases 1 to 3173)
  AUTHORS   Kirisawa,R., Kobayashi,T., Kawakami,Y. and Iwai,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-1996) to the DDBJ/EMBL/GenBank databases. Rikio
            Kirisawa, Rakuno Gakuen University, Faculty of Veterinary Medicine,
            Department of Veterinary Microbiology; Bunkyodai-Midorimachi 582,
            Ebetsu, Hokkaido 069, Japan (E-mail:r-kirisa@ca2.so-net.or.jp,
            Tel:+81-11-386-1112, Fax:+81-11-387-5890)
REFERENCE   2  (sites)
  AUTHORS   Kirisawa,R., Kobayashi,T., Kawakami,Y. and Iwai,H.
  TITLE     Nucleotide sequences of open reading frames 1, 24 and 71 of an
            attenuated equine herpesvirus-1
  JOURNAL   J. Equine Sci. 7, 79-87 (1996)
FEATURES             Location/Qualifiers
     source          1. .3173
                     /organism="Equine herpesvirus 1"
                     /isolate="3F clone"
                     /strain="BK343"
                     /db_xref="taxon:10326"
     gene            118. .2721
                     /gene="ORF71"
     CDS             118. .2721
                     /gene="ORF71"
                     /codon_start=1
                     /product="membrane glycoprotein"
                     /db_xref="PID:d1020846"
                     /db_xref="PID:g2114323"
                     /translation="MGFIYARKLLLCMAVSIYAIGSTTTTETTTSSSSTSGSGQSTSS
                     GTTNSSSSPTTSPPTTSSSPPTSTHTSSPSSTSTQSSSTAATSSSAPSTASSTTSIPT
                     STSTETTTTTPTASTTTPTTTTAAPTTAATTTAVTTAASTSAETTTATATATSTPTTT
                     TPTSTTTTTATTTVPTTASTTTDTTTAATTTAATTTAATTTAATTTAATTTAATTTAA
                     TTTAATTTAATTTAATTTAATTTAATTTAATTTAATTSSATTAATTTAATTTAATTTA
                     ATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTGSPT
                     SGSTSTTGASTSTPSASTATSATPTSTSTSAAATTSTPTPTSAATSAESTTEAPTSTP
                     TTDTTTPSEATTATTSPESTTVSASTTSATTTAFTTESHTSPDSSTGSTSTAEPSSTF
                     TLTPSTATPSTDQFTGSSASTESDSTDSSTVPTTGTESITESSSTTEASTNLGSSTYE
                     STEALETPDGNTTSGNTTPSPSPRTPSFADTQQTPDNGVSTQHTTINDHTTANAQKHA
                     GYHRGRAGGRRGIPQGGSHTTPHPDRLTPSPDDTYDDDTNHPNGRNNSIEIVPQLPPD
                     RPIIELGVATLRKNFMEASCTVETNSGLAIFWKIGNASVDAFNRGTTHTRLMRNGVPV
                     YALVSTLRVPWLNVIPLTKITCAACPTNLVAGDGVDLNSCTTKSTTIPCPGQQRTHIF
                     FSAKGDRAVCITSELVSQPTITWSVGSDRLRNDGFSQTWYGIQPGVCGILRSEVRIHR
                     TTWRFGSTSKDYLCEVSASDSKTSDYKVLPNAHSTSNFALVAATTLTVTILCLLCCLY
                     CMLTRPRASVY"
     repeat_region   667. .861
     repeat_region   892. .1101
     polyA_signal    2802. .2807
BASE COUNT      857 a   1068 c    643 g    605 t
ORIGIN
```

Query Match 5.7%; Score 140; DB 17; Length 3173;
Best Local Similarity 51.1%; Pred. No. 3.8e-17;
Matches 329; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

```
Qy   1492 CACATCGACTGCTGCCCCGACAGCTGCTCCAACGGCTGCCCCAACGGCTGCCCCAACGGC 1551
          | |||| ||  |  |||| ||| |  |  |  ||||| || |||||| ||  |||| |
Db    459 CGCATCTACAACGACCCCAACAACAACGACCGCGGCTCCCACAACGGCCGCTACAACCAC 518

Qy   1552 TGCCCCAACTGCCGCACCCTCCACCGTGGTCCCACCTGCAACGCCACCCGCAACTGCAGC 1611
           ||    ||| | ||  | || ||    |    | |  | ||  |  | || |||||  |
Db    519 AGCTGTTACTACAGCCGCGTCTACATCAGCGGAAACCACCACAGCTACTGCGACTGCTAC 578

Qy   1612 CCCAGTCCCACCTACAACCGCAATTCCTACTCCGGCCCCCACCGCTGCCCCCACCGCAGC 1671
          | ||  |||| | |||||  |   | |  |  |  |  | ||  | ||  ||||| | |
Db    579 CTCAACCCCAACCACAACTACGCCTACGTCCACAACAACTACTACAGCTACCACCACTGT 638

Qy   1672 TCCTACTACTGCTGCCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCC 1731
          |||  ||  ||  |||  |  |    | |    || ||     ||| | |  | || ||  |
Db    639 TCCAACAACCGCTTCTACAACAACTGATACGACCACAGCAGCAACGACCACAGCAGCAAC 698

Qy   1732 CACCGCAGCCCCTACTACTGCTGTCCCTGAAATCCCAATCACTGTCACATCAGCGCCTAC 1791
           ||| ||||  | || ||  | |   |    | | ||    |    ||  ||||  | ||
Db    699 GACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAAC 758

Qy   1792 CGCTGCCCCCACCGCTGCCCCCACCGCTGCCCCCACCGCAGCCCCTACTACTGCTGTCCC 1851
            |  |  |  |  |  || |  | ||  |  |||| |||||  | || || || |   |
Db    759 GACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAAC 818

Qy   1852 AGAAATCCCAACTACTGTCACATCACCACCTACTGCTGCCCCCACTACCGCAGCACCTGC 1911
            |   || | ||   |||| || || | ||  | ||  | || ||| | ||  || |
Db    819 GACCACAGCAGCAACGACCACAGCAGCAACGACCACAGCAGCAACGACCACCGCGGCTAC 878

Qy   1912 CCCCAACACCACAGTCACTGTACCACCCACTGCTGCCCCTACTACCGCAGCACCTGCCCC 1971
            |   | |  ||  ||| |   |  |||| | ||  |||| ||| | ||  || || |
Db    879 TACTTCCTCTGCAACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCAC 938

Qy   1972 CAACACCACAGTCACTGTACCACCCACTGCTGCCCCCACTGCAGCTCCCCCTACCGTCGC 2031
          || | |  |   |||      |  | ||  |  || |  || |  |  || |  |   | |
Db    939 CACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCAC 998

Qy   2032 ACATGCACCCAACACCACAGCTGCCCCGGTAACTACAACCAGCGCACCAGCTACCACACC 2091
            ||  | | ||||||| || ||  |     ||  |  | | | |  | ||  | ||  |
Db    999 CACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCAC 1058

Qy   2092 TGAAGATGATGACATCGACCCCCCTCTCCCCAACGACCCCATCA 2135
              |  | |  || |  | |  ||  | ||| ||   | | ||
Db   1059 CACCGCGGCTACCACCACCGCGGCTACCACCACCGCGGCTACCA 1102
```

Search completed: July 18, 1999, 05:58:19
Job time: 2985 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 00:46:11 ; Search time 26.5 Seconds
(without alignments)
615.918 Million cell updates/sec

Title: US-09-294-663-4
Perfect score: 4395
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 807

Scoring table: BLOSUM62

Searched: 162890 seqs, 20225328 residues

Database : A_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	542	12.3	1404	1	R26049	MSF precursor. New
2	453.5	10.3	1721	1	W48299	Cryptosporidium pa
3	383	8.7	763	1	W31852	Mycobacterium tube
4	372.5	8.5	1664	1	W43106	C. thermocellum Ol
5	335.5	7.6	572	1	W31855	Mycobacterium tube
6	332.5	7.6	297	1	W15761	Cotton fibrous tis
7	331	7.5	400	1	P91941	Sequence of prepro
8	327	7.4	844	1	P60570	Sequence of the Fa
9	325	7.4	907	1	P50073	Epstein-Barr virus
10	325	7.4	907	1	R80144	EBV gp350/220. New
11	313.5	7.1	278	1	R07671	Intestinal mucin d
12	309.5	7.0	267	1	W64359	Mycobacterium tube
13	309.5	7.0	267	1	W81726	M. tuberculosis im
14	307.5	7.0	451	1	R92803	Hepatitis A virus
15	303	6.9	278	1	R12535	SMUC-41 intestinal
16	282	6.4	166	1	R07670	Intestinal mucin d
17	280	6.4	166	1	R12534	SMUC-40 intestinal
18	276.5	6.3	214	1	R86913	Cotton fiber-speci
19	254.5	5.8	452	1	R80041	Human megakaryocyt
20	252.5	5.7	575	1	R11862	Delta-endotoxin Pv
21	248.5	5.7	3119	1	W72204	HSV-2 strain SB5 C
22	247	5.6	1537	1	R60562	Yeast 4.7 kB agglu
23	247	5.6	849	1	W06725	FLO1 protein, invo
24	239	5.4	486	1	R08034	Unique sequence fr
25	237.5	5.4	895	1	R40929	43 kDa/156 kDa non
26	237.5	5.4	895	1	R80230	Rabbit 97 kDa dyst
27	237.5	5.4	895	1	W31588	Rabbit (156 kDa) d
28	236.5	5.4	895	1	R80231	Human dystroglycan
29	236.5	5.4	895	1	W31589	Human dystrophin-a
30	233	5.3	334	1	P82971	Bioadhesive precur
31	232.5	5.3	2035	1	R57141	Host cell factor p
32	230.5	5.2	895	1	R50240	Sequence of precur
33	229.5	5.2	652	1	P82974	Bioadhesive precur
34	229.5	5.2	733	1	W69430	Human secreted pro
35	229	5.2	334	1	W38336	Human kidney injur
36	228	5.2	334	1	P83194	Sequence of a bioa
37	227.5	5.2	744	1	P82975	Bioadhesive precur
38	227	5.2	841	1	W34985	Archaebacterial th
39	224	5.1	357	1	R66810	Extracellular doma
40	223	5.1	783	1	W37151	Mouse neural Mena+
41	223	5.1	787	1	W37152	Mouse neural Mena+
42	223	5.1	802	1	W37153	Mouse neural Mena+
43	223	5.1	508	1	W77233	MiniMUC1 protein.
44	219.5	5.0	629	1	R88466	Drosophila scaveng
45	217.5	4.9	132	1	R07673	Intestinal mucin d

ALIGNMENTS

RESULT 1
R26049
ID R26049 standard; Protein; 1404 AA.
AC R26049;
DT 02-FEB-1993 (first entry)
DE MSF precursor.
KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
KW stability; proteolytic cleavage; adhesion; alternative splicing.
OS Synthetic.
FH Key Location/Qualifiers
FT region 1. .26
FT /label= Exon_I
FT region 26. .67
FT /label= Exon_II
FT region 67. .107
FT /label= Exon_III
FT region 107. .157
FT /label= Exon_IV
FT region 157. .200
FT /label= Exon_V
FT region 200. .1141
FT /label= Exon_VI
FT region 1411. .1166
FT /label= Exon_VII
FT region 1166. .1212
FT /label= Exon_VIII
FT region 1213. .1266
FT /label= Exon_IX
FT region 1266. .1331
FT /label= Exon_X
FT region 1331. .1373
FT /label= Exon_XI
FT region 1373. .1404
FT /label= Exon_XII
PN WO9213075-A.
PD 06-AUG-1992.
PF 17-JAN-1992; U00433.
PR 18-JAN-1991; US-643502.
PR 10-SEP-1991; US-757022.
PA (GEMY) GENETICS INST INC.
PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
DR WPI; 92-284660/34.
DR N-PSDB; Q27223.
PT New human mega-karyocyte stimulating factors - for treating
PT immune deficiencies, cancer, exposure to radiation or drugs,
PT bacterial and viral infections, etc.
PS Claim 1, 2 and 3; Fig 1; 87pp; English.
CC The sequence given is a full length translation from the megakaryocyte
CC stimulating factor (MSF) precursor. The sequence covered by exons II,
CC III and IV encodes megakaryocyte stimulating factor (MSF). This
CC sequence is modified by the addition of an N-terminal sequence encoding
CC a secretory leader, an initiating methionine proceeding exon II and a
CC terminating codon following exon IV. The cDNA sequence given contains
CC sequences derived from human megakaryocyte colony stimulating factor
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a
CC classical mammalian protein secretion signal sequence. The sequence
CC encoding the original meg-CSF includes exons II-IV and is thought to
CC terminate in the region between amino acid residues 134 - 147. The
CC primary transcript of this gene may be cleaved in different ways to
CC yield a family of mRNA's each encoding a different MSF protein. Exons
CC V and VI are thought to be related to the activity of the factor and
CC are also implicated in the stability, folding and processing of the
CC molecule. These exons are also thought to play a role in the observed
CC synergy of MSF with other cytokines. Exons V - XII are believed to be
CC implicated in the processing or folding of the appropriate structure of
CC the resulting factor, ie. one or more of these exons may contain

CC sequences which direct proteolytic cleavage, adhesion, organisation of
CC the cellular matrix or extracellular matrix processing. Both naturally
CC occuring and non-naturally occuring MSF's may be characterised by
CC various combinations of alternatively spliced exons from this sequence,
CC with the exons spliced together in differing orders to form different
CC members of the MSF family.
SQ Sequence 1404 AA;

Query Match 12.3%; Score 542; DB 1; Length 1404;
Best Local Similarity 30.4%; Pred. No. 6.6e-22;
Matches 215; Conservative 23; Mismatches 194; Indels 276; Gaps 37;

```
Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142
           | | || ||: || |      ||| : | :||     |||   ||||  : |||||: ||
Db     331 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 390

Qy     143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192
           || :  |||||: |  ||    ||||| : |||||  |||| :  |||| :  |||| :
Db     391 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 450

Qy     193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
           |     | |||   ||  |||    |    || |||  |||| ||  |
Db     451 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 492

Qy     253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
                                                |   : :|   :      |||
Db     492 ------------------------------------PTTPKEPAPTTPK-----EPAPT 509

Qy     313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372
           |:                 |:
Db     510 TTKE---------------PS----------------------------------- 516

Qy     373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429
                                              || |   || |      |
Db     516 ----------------------------------PTTP--KEPAPTTTKSAPT------ 533

Qy     430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489
                         |   | |     :| |    || |               |
Db     533 ------------------TTKEPAPTTTKSAPTTPKEPSPTT-------------TKEP 560

Qy     490 APTA----APTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
           |||     |||     |||     ||:|   || |  |  | |  ||   ||||     |
Db     561 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 617

Qy     546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593
           |||    :|||      | |   |||     |||| || |   |    |||   |||
Db     618 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 675

Qy     594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT---------- 629
                |||||     |||      |:|   ||||           ||| |
Db     676 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 735

Qy     629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683
           | |  ||:  |:  |||     |  | |   ||||   ||| |     ||     || |
Db     736 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 794

Qy     684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720
                 || ||  | :|||   || |  |     |  | |  |  |
Db     795 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 842
```

RESULT 2
W48299
ID W48299 standard; Protein; 1721 AA.
AC W48299;
DT 17-AUG-1998 (first entry)
DE Cryptosporidium parvum GP900 antigen.
KW Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW antibody; prophylaxis; treatment; inhibition; retardation;
KW detection; diagnosis; human; 3' region.
OS Cryptosporidium parvum.
FH Key Location/Qualifiers
FT Domain 175. .423
FT /note= "NINC domain"
PN WO9806430-A1.
PD 19-FEB-1998.
PF 11-AUG-1997; U14104.
PR 14-AUG-1996; US-700651.
PA (REGC) UNIV CALIFORNIA.
PI Gut J, Leech J, Nelson RC, Petersen C;
DR WPI; 98-159290/14.
DR N-PSDB; V20700, V20701.
PT Anti-Cryptosporidium antibody - used to develop products for
PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT infections
PS Claim 27; Pages 65-68; 89pp; English.
CC The sequence is that of the GP900 antigen which may be used
CC in the production of anti-Cryptosporidium antibodies. These can be
CC used for the prophylaxis, treatment, inhibition or retardation of
CC a Cryptosporidium infection in humans or in animals such as calves.
CC They can also be used for the detection and diagnosis of related
CC infections.
SQ Sequence 1721 AA;

Query Match 10.3%; Score 453.5; DB 1; Length 1721;
Best Local Similarity 27.1%; Pred. No. 4.4e-17;
Matches 175; Conservative 40; Mismatches 246; Indels 185; Gaps 20;

```
Qy      99 TTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTT 158
           ||    ||   |||   ||||   ||||   ||||: ||||   |||   |||   ||||
Db     270 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTT 329

Qy     159 QAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATT-PAATTPAATT 217
              ||||: |||||   |||   |||   |||   || |   ||   || |  ||   ||
Db     330 TTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKPTTTTTTTT 389

Qy     218 PAATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYT 277
                 |  ||       |::                  :   :|  |::| |
Db     390 TTTKKPTTTTTATTTTTTS-----------------ETESVIKPDEWCWL--------- 423

Qy     278 FEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFE 337
                                  .  | ||: : :                :| :  :|   |
Db     423 -------------------EKNGECEAKGATY------------VGVIGKDGRIENGM 449

Qy     338 IDWLLPHGNRCDKYYQCVHGNLVERRCGAGTH--FSFELQQCDHIELVGCTLPGGESEEV 395
              ::|: :                      ||  | |:::   :   | |    |: |
Db     450 AFTMIPNDD--------------------THVRFRFKVKDVGNTISVRCRKGAGKLEFP 488

Qy     396 DVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPG 455
           |                              ||:|  :  | | |   :  | |       |
Db     489 DRS-----------------------LDFTIPPVAGHNS-C-SIIVGVSGD-------G 515

Qy     456 NLHFSPATQSCESPVTAGCQVFECDSDNQC-TSTAAPTAAPTAAPTAAPTAAPSTV 514
            :| ||      | ::|  |  |  ::  | | ||   |  :   |:|     | |  |
Db     516 KIHVSPYGSKDVSLISAPIQPSELFNEVYCDTCTAKYGAIHSGYQTSADFVTTTTAKPTT 575

Qy     515 VPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVP 574
                 | ||  |||                || :|  |||  |       | | ||
Db     576 T------TTGAPGQPTTT-------------TTGSPSKPTTTTT------TKATTTTTT 609

Qy     575 EIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAP---TTAAPAPNTTVT 631
             ||  |:           |  ||   || ||  |   |:  |  |   ||   |  || |
Db     610 LNPIITTT----------TQKPTTTTTTKVPGKPPIATTTTTLKPIVTTTTTKATTTTTT 659

Qy     632 VPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAP--------------NTTVTVPPTAAP 676
             ||   ||       || |  |         | |                |: : :     |
Db     660 TVPTTTTTTKRDEMTTTTPLPDIGDIEITPIPIEKMLDKYTRMIYDYNSGLLLDSNDEP 719

Qy     677 TAAPPTVAHAPNTTAAPVTT---TSAPATTPEDDDIDP--PLPNDP 717
                  |  :   || |   |  |         |||   || ||
Db     720 IPGSQAGQIADTSNLFPVQTHKSTGLP--------IDPMVGLPFDP 757
```

RESULT 3

```
W31852
ID   W31852 standard; Protein; 763 AA.
AC   W31852;
DT   27-APR-1998  (first entry)
DE   Mycobacterium tuberculosis 74 kDa protein.
KW   Tuberculosis; mycobacteria; infection; diagnosis;
KW   antimycobacterial; antibiotic; vaccine.
OS   Mycobacterium tuberculosis.
PN   WO9741252-A2.
PD   06-NOV-1997.
PF   18-APR-1997; E01973.
PR   29-APR-1996; DE-017184.
PA   (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
PI   Espitia C, Honisch C, Moreno C, Singh M;
DR   WPI; 97-549750/50.
DR   N-PSDB; T93610.
PT   New DNA and related proteins or RNA derived from M. tuberculosis -
PT   used for diagnosis of mycobacterial infections, monitoring
PT   vaccination and development of anti-mycobacterial agents
PS   Claim 5; Fig 13; 55pp; English.
CC   This novel 74 kDa protein is encoded by an open reading frame of
CC   a Mycobacterium tuberulosis DNA fragment (see T93610) containing
CC   polymorphic GC-rich sequences.  Its amino acid sequence shows
CC   a high proline content, but there is no homology to any known
CC   proline-rich antigens of mycobacteria.  Novel M. tuberculosis
CC   proteins (see W31851-57) are claimed.  These can be produced as
CC   recombinant proteins, especially in bacterial, yeast, fungal or
CC   higher eukaryote host cells, and used for diagnosing tuberculosis
CC   and other mycobacterial infections in humans or animals.  The
CC   claimed proteins can also be used for epidemiological studies, for
CC   monitoring vaccination, and for the development of vaccines and
CC   anti-mycobacterial drugs.
SQ   Sequence   763 AA;

  Query Match             8.7%;  Score 383;  DB 1;  Length 763;
  Best Local Similarity   23.2%;  Pred. No. 1.2e-13;
  Matches  194;  Conservative   40;  Mismatches  299;  Indels  302;  Gaps   35;

Qy      14 VAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRDCAPGT 73
           :|  |    |   |   | |  |||    :|                        |   | |
Db      10 LAPLPPAPPAPAEPKSKPPFPPAPPAPPCWMLVS----------------AAPPCPPAP 52

Qy      74 EFKFSAQTCVHAALAGCTLPGPPAETTQA---------PATTQAPTTTQAPTTTTQAPTT 124
                          | ||   ::|          ||   ||    ||    :||
Db      53 -------------------PAPPKPKSKAPFPPVPPAPPARELAPPLPPAP---PEAPRE 90

Qy     125 TTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTT 184
           :  |       |    |    ||   || :   |                 ||    ||
Db      91 SRPALPPCPPPPVVIPDPPEPAAPPVPPAPNSPPFPP---------FPPAPKFVPAPPVP 141

Qy     185 QAPTTTQAPTTITQAATTPAATTPAATT-----PAATTPAATTPAATTPGVP-APTSAPV 238
             | :   |    |   ||   |          ||  ||| | |||  | || || | |
Db     142 PVPNSPPFPPFPPAALNPPAPPAPPLANSPPLPPAPPTPAGTPPAAPWPPVPAAPKSKPA 201

Qy     239 WPPICELLP-NGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDS 297
            ||      |    | :|     :| |                :  |       |
Db     202 SPPRPPAPPMPATPMEFPPLPPVPPD------------PISKETPPAPPAPPI-----P 243

Qy     298 PANV---------ECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRC 348
           || |            :| ||||         : :  :|   ||        ||
Db     244 PAPVPIPPVPPLPPVPNKIPPAPPA------PPVAVAAVLVAPCPP-------LP----- 286

Qy     349 DKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP 408
                                                 ||                |
Db     286 ------------------------------------PLPNNHPPAPPA---------AP 299

Qy     409 TEPIEWEPLPNG---CPADFSIDHL----LPHESDCGQYLQCVHGQTIA---------- 451
             :   ||||     |    : :    ||           : |: ::
Db     300 VPGVPLAPLPNSHPPAPPSAPVPGVPLAPLP----------ISGRPVSVWKGSFTTLST 348

Qy     451 ---RPCPGNL---HFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAP--------- 496
               | | | :     :|:  |  ||:|              |: | |   |

Db     349 FCCRVCSGEVLAGALNPSRPS-RSPLTT------------TTPALPAPIPPLPPLPPLP 394

Qy     496 --TAAP--------TAAPTAAPTAAP---STVVPPA----------TPPATAAP------ 527
             || |         ||      |  ||    |  ||||           |||   ||
Db     395 INTAVPPIPPLPPVTALAPPLPPLAPLPISPGVPPAPPIPPGKPWTTPPLAPAPPEPKTV 454

Qy     527 --VPPTTAIP------------TPAPTAAPTAAPTTAAPESPTTVTVPPT----AAPTAA 568
             :||  : |              | | ::|   |   ||  |: | |||:     || ||
Db     455 PVLPPGPSCPPSEKPNPPAPPEPPEPKSSPALPPAPPAPSMPSAVRVPPSPPIPPAPPAA 514

Qy     569 PTTAVPEIPITVTSAPTA----------------APTAAPT------AAPTAAPTTAVP 605
           |  ::| :|    | |                  || | ||      | |   |    |
Db     515 PRASMPALPPAPPSPPATRLCPPLPPSPPAPNSPPAPPAPPTPPKLLSANPPCPPVPPAP 574

Qy     606 EIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPN 665
             |    :||       | | | |    || |    | |||     |||:| |    |||
Db     575 NRPPAPPAPPAPPELPAPPDPPT----PPVANSPPAPPAP----PAPPSALPFVNPPAP- 626

Qy     666 TTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 720
                 |  ||| : |    |  |    | ||  |: |    |      || ||    |
Db     626 -----PTPAAPKSRPALPAAPPAPPAPPVRATTPPPAPPA-----PPAPNSMALP 670


RESULT   4
W43106
ID   W43106 standard; Protein; 1664 AA.
AC   W43106;
DT   16-OCT-1998  (first entry)
DE   C. thermocellum OlpB protein.
KW   Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
KW   cohesin domain; catalytic subunit; scaffold subunit; SdbA; synergistic;
KW   cellulosome integrating protein; scaffoldin dockerin binding protein.
OS   Clostridium thermocellum.
FH   Key             Location/Qualifiers
FT   Domain          28. .192
FT                   /note= "cohesin type II domain"
FT   Domain          207. .363
FT                   /note= "cohesin type II domain"
FT   Domain          409. .565
FT                   /note= "cohesin type II domain"
FT   Domain          607. .763
FT                   /note= "cohesin type II domain"
PN   FR2748479-A1.
PD   14-NOV-1997.
PF   10-MAY-1996; 005854.
PR   10-MAY-1996; FR-005854.
PA   (INSP ) INST PASTEUR.
PI   Beguin P, Leibovitz E;
DR   WPI; 98-011569/02.
DR   N-PSDB; T86623.
PT   Cellulase proteins with cohesin or dockerin type II domains - useful
PT   for potentiating the activity of multiprotein enzyme complexes
PS   Claim 7; Page 31-39; 60pp; French.
CC   Multimeric protein, especially enzymatic, complexes are held together
CC   by protein-protein interactions between domains designated dockerins
CC   and cohesins, which are found on the catalytic and scaffold subunits
CC   respectively.  An example of such a complex is the cellulose degrading
CC   protein complex from Clostridium thermocellum, known as the cellulosome.
CC   This complex comprises around 15 proteins including endoglucanases,
CC   cellobiohydratases, hemicellulases, e.g. xylanases or lichinases, which
CC   interact with a central "scaffold" protein designated the cellulosome
CC   integrating protein (CipA; see W43108).  The catalytic subunits interact
CC   with the Cip subunit via conserved 23 amino acid dockerin domains. Cip
CC   has been shown to contain 9 copies of a cohesin domain.
CC   The invention relates to the isolation of proteins binding to a novel
CC   dockerin type domain found in the C-terminal portion of Cip.  The new
CC   domain is designated a type II dockerin domain (as compared to the type I
CC   domain found on the catalytic subunits of the cellulosome).  The type II
CC   dockerin domain has some sequence similarity to the type I dockerins but
CC   is unable to bind type I cohesin domains.
CC   The sequence presented here is an example of a protein which binds
CC   the novel type II dockerin domain and is the product of the OlpB gene.
CC   The protein contains 4 type II cohesin domains in the N-terminal portion
```

```
CC   of which the first domain (amino acid residues 28-192) is thought to bind
CC   CipA.  The novel type II dockerin and cohesin domains can be used in
CC   complexes, especially enzyme complexes, to potentiate their catalytic
CC   actions in a synergistic manner.
SQ   Sequence   1664 AA;


  Query Match             8.5%;  Score 372.5;  DB 1;  Length 1664;
  Best Local Similarity   23.5%;  Pred. No. 9.1e-13;
  Matches  176;  Conservative   66;  Mismatches  325;  Indels  183;  Gaps   30;

Qy      18 PEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRD-CAPGTEFK 76
           |  || |  |:  || |   |       | |         |  :    | |   |  |
Db     825 PTPSD-EPTPS-DEPTPSDEPTPSDEPTPSETP-------EEPIPTDTPSDEPTPSDE-- 874

Qy      77 FSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAP 136
                     | |  | | :   | :   | : |     || |     | :  |
Db     874 ----------------PTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEP 917

Qy     137 TTTQAPTTTQAPTTTQAPTTT----TQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQA 192
           | :  || :  || :  || :       || |     | :  || :  || :  || :
Db     918 TPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 977

Qy     193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPG-VPAPTSAPVWPPICELLPNGCP 251
           ||   :   ||:     : ||    |  |     ||   | |:  |   |  |  |:  |
Db     978 PTPSDE--PTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPT--PSDEPTPSDEP 1033

Qy     252 ADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAP 311
             |    | |       : :   | |:  |          |:|::    | :|:
Db    1034 TPSDEP--TPSD-------EPTPSETPEEPIP----------TDTPSD-----EPTPSD 1068

Qy     312 PVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFS 371
             |  :|    |     |   |:    :  :|     |
Db    1069 EPTPSDEPTPSDEPTPSDEPTPSETP-EEPIPTDTPSD---------------------- 1106

Qy     372 FELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLL 431
                |      |    |    ||   :    |:|  | ||:|   |:|
Db    1106 ---------EPTPSDEPTPSDEPTPSDEPTPSDEPTPSETPE-EPIPTDTPSD------E 1149

Qy     432 PHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAP 491
           |  ||                | | :   |      :|          ||    |
Db    1150 PTPSD-------------EPTPSD---EPTPSDEPTP----------SDEPTPSETPE 1181

Qy     492 TAAPTAAPTAAPTAAPTAAPSTVVPPATPPATA-APVP---PTTAIPTPAPTAAPTAAPT 547
             ||  |:  || :    ||    |:  |  :  | |   |   |||  |:  || :
Db    1182 EPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDE 1241

Qy     548 TAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPE 606
             : ||    | |:  || : |   || ||  |  |:  || :    |:  || :
Db    1242 PTPSDEPTPSDEPTPSDEPTPSET---PEEPIP-TDTPSDEPTPSDEPTPSDEPTPSDEP 1297

Qy     607 IPTTVTSP---------PTAAPTTAAPAPNTTVTVPPTAA-------------PTTA-- 642
             |:   :|         || : |   | |  | :  || :              || :
Db    1298 TPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1357

Qy     642 -APAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAP-------------PTVAHA 686
             |:   | :  || : |   | | || |  |:  ||:                ||  :
Db    1358 PTPSDEPTPSDEPTPSETPEEPTPTTTPTPTPSTTPTSGSGGSGGSGGGGGGGGGGTVPTS 1417

Qy     687 PNTTAAPVTTTSAPATTPEDDDIDPPLPND 716
           |       | || | :||   :|: | |:|
Db    1418 PTP-----TPTSKPTSTPAPTEIEEPTPSD 1442


RESULT   5
W31855
ID   W31855 standard; Protein; 572 AA.
AC   W31855;
DT   27-APR-1998  (first entry)
DE   Mycobacterium tuberculosis 55 kDa protein.
KW   Tuberculosis; mycobacteria; infection; diagnosis;
KW   antimycobacterial; antibiotic; vaccine.
OS   Mycobacterium tuberculosis.
PN   WO9741252-A2.
PD   06-NOV-1997.
PF   18-APR-1997; E01973.
PR   29-APR-1996; DE-017184.
PA   (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
PI   Espitia C, Honisch C, Moreno C, Singh M;
DR   WPI; 97-549750/50.
DR   N-PSDB; T93610.
PT   New DNA and related proteins or RNA derived from M. tuberculosis -
PT   used for diagnosis of mycobacterial infections, monitoring
PT   vaccination and development of anti-mycobacterial agents
PS   Claim 11; Fig 16; 55pp; English.
CC   This novel 55 kDa protein is encoded by an open reading frame of
CC   a Mycobacterium tuberculosis DNA fragment (see T93610) containing
CC   polymorphic GC-rich sequences.  Its amino acid sequence shows
CC   a high proline content, but there is no homology to any known
CC   proline-rich antigens of mycobacteria.  Novel M. tuberculosis
CC   proteins (see W31851-57) are claimed.  These can be produced as
CC   recombinant proteins, especially in bacterial, yeast, fungal or
CC   higher eukaryote host cells, and used for diagnosing tuberculosis
CC   and other mycobacterial infections in humans or animals.  The
CC   claimed proteins can also be used for epidemiological studies, for
CC   monitoring vaccination, and for the development of vaccines and
CC   anti-mycobacterial drugs.
SQ   Sequence   572 AA;


  Query Match             7.6%;  Score 335.5;  DB 1;  Length 572;
  Best Local Similarity   22.7%;  Pred. No. 3e-11;
  Matches  170;  Conservative   41;  Mismatches  282;  Indels  257;  Gaps   32;

Qy      20 VSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSA 79
             |  | |:     | |  ||                                | |  :|
Db       1 VPAAPKSKPASPPRPPAPP-----------------------------MPATPMEFPP 29

Qy      80 QTCVHA-ALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTT 138
             |    ::  | | |||     ||    |     |    : |      |
Db      30 LPPVPPDPISKETPPAPPAPPI-PPAPVPIPPVPPLPPVPNKIPPAPPAPPVAVA---AV 85

Qy     139 TQAPTTTQAPTTTQAPTTTTQAP---TTTTQAPTTTTQAPTTTQAPTTTQAPTTTQA-PT 194
             ||     |    |    ||        | :  || :  |   ||     |
Db      86 LVAPCPPLPPLPNNHPPAPPAAPVPGVPLAPLPNSHPPAPPSAPVPGVPLAPLPISGRPV 145

Qy     195 TITQAATTPAAT-----------------TPAATTPAATTPAATTPAATTPGVP----- 232
           :: : : |  :|                  |: :    ||||   |    | :|
Db     146 SVWKGSFTTLSTFCCRVCSGEVLAGALNPSRPSRSPLTTTTPALPAPIPPLPPLPPLPIN 205

Qy     232 -APTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNP 290
            |    |  ||:  | |   |       | |           | |       | |    |
Db     206 TAVPPIPPLPPVTALAPPLPP-------LAP--------LPISPGVPPAPPIPPG---KP 247

Qy     291 YVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDK 350
           :    :|       ::||||     | : : :                  || | |
Db     248 WT----TP-------PLAPAPP-----EPKTVPV----------------LPPGPSC-- 273

Qy     351 YYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTE 410
                                          | |                 | |
Db     273 -------------------------------------PPSEKPNPP----------APPE 285

Qy     411 PIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPV 470
           |    | |   ||       ||                  | |       | :   ||
Db     286 P----PEPKSSPA-------LP------------------PAPPAPSMPSAVRVPPSP- 315

Qy     471 TAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPST----VVPPATPPATAAP 526
                             || ||| |:  | |  || : |:|     :||: |   : |
Db     315 ---------------PIPPAPPAAPRASMPALP-PAPPSPPATRLCPPLPPSPPAPNSPP 358

Qy     527 VPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTS---- 583
            ||    |    :| |   |   ||  |     ||  ||   |    |  |    |
Db     359 APPAPPTPPKLLSANPPCPPVPPAPNRPPAPPAPP--APPELPAPPDPPTPPVANSPPAP 416

Qy     583 -APTAAPTAAPTAAPTAAPTTAVPEI-PTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTT 640
            || | |:| |   | | || | |:  |    :||      | |||    | || | |
```

```
Db    417 PAPPAPPSALPFVNPPAPPTPAAPKSRPALPAAPP------APPAPPVRATTPPPAPP-- 469

Qy    641 AAPAPNTTVTAP-------PTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAP 693
          | ||||:    |       |  |   | |||   :: |       |||    ||:  | |
Db    469 APPAPNSMALPPAPPDPPIPLLATPPAPPAPPLPMSPP-------APPLPPAAPDPPAPP 521

Qy    694 VTTTSAPATTPEDDDIDPPL---PNDPINP 720
          :|    |:          |||   |  |: |
Db    522 LTINQPPS---------PPLAPVPGAPLAP 542


RESULT   6
W15761
ID   W15761 standard; Protein; 297 AA.
AC   W15761;
DT   19-JUN-1997  (first entry)
DE   Cotton fibrous tissue specific protein KC03.
KW   Fibrous tissue; cotton fibre; cotton flower; Gossypium barbadense;
KW   Gossypium hirsutum.
OS   Gossypium barbadense.
PN   J09075093-A.
PD   25-MAR-1997.
PF   20-FEB-1996; 031987.
PR   21-FEB-1995; US-391696.
PR   29-DEC-1995; US-580545.
PA   (TOYM ) TOYOBO KK.
PA   (UYTE-) UNIV TEXAS TECH.
DR   WPI; 97-239274/22.
DR   N-PSDB; T60569.
PT   Cotton fibrous tissue gene - used to produce transformants with
PT   improved cotton fibres, and improved yield
PS   Claim 3; Page 21-22; 37pp; Japanese.
CC   W15759-W15763 represent cotton fibrous tissue proteins of the invention.
CC   The genes encoding these sequences are specifically expressed during the
CC   growth of cotton fibre. The genes were isolated from the polyA RNA of the
CC   ovules of a cotton flower of Gossypium barbadense, and G. hirsutum. The
CC   genes encoding these sequences, and vectors containing them are used in
CC   the preparation of cotton transformants, with improved cotton fibres and
CC   improved yield.
SQ   Sequence   297 AA;


  Query Match             7.6%;  Score 332.5;  DB 1;  Length 297;
  Best Local Similarity   36.4%;  Pred. No. 2.3e-11;
  Matches  103;  Conservative    8;  Mismatches  115;  Indels   57;  Gaps   16;

Qy    497 AAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTT 556
          |||   |  ||  |     ||  || |  ||||  | |||||   ||  ||   |  ||
Db     35 AAP--VPVKAPIPA-----PPVKPPTT--PVPPYKA-PTPAP---PTKGPTPYKP--PTK 79

Qy    557 VTVPPTAAPTAAPTTAV-PEIPITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPP 615
            ||   || |||    |  |   | |||        | || |||   |      :||
Db     80 APTPPYKPPTKAPTPPYKPPAPAPPTKAPTPPYKPPAPAPPTKAPT---PPYKPPAPAPP 136

Qy    616 TAAPT----TAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVP 671
          | |||      |||| |    ||   | | ||||      ||   ||    || |    |
Db    137 TKAPTPPFKPPAPAPPTKAPTPP-YKPPTPAPAPPVKAPTPPYMPPTPPTKAP-TPAPAP 194

Qy    672 PTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWA 731
          || |||  ||     |     | || : |   |      ||||  |:
Db    195 PTKAPT--PPYKPPVPTPPVKPPTTPAPPYKPP-----SPPLP--PVR----------T 234

Qy    732 HADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVR 774
            ||          : :| :  : :  |  |  ||   | |
Db    235 KKDC-----------IPLCGQRCKLHSRTNLCLRACMTCCDR 265


RESULT   7
P91941
ID   P91941 standard; peptide; 400 AA.
AC   P91941;
DT   07-FEB-1990  (first entry)
DE   Sequence of preprospasmolysin.
KW   Preprospasmolysin; Xenopus laevis.
OS   Xenopus laevis
PN   DE3808456-A.
PD   28-SEP-1989.
PF   14-MAR-1988; 808456.
PR   14-MAR-1988; DE-808456.
PA   (PLAC) Max-Plank Ges. Wissensch.
PI   Weickmann H, Fincke K, Weickmann F, Huber B, Liska H, Prechtel J;
DR   WPI; 89-286097/40.
DR   N-PSDB; N91235
PT   Polypeptide preprospasmolysin and its fragments - derived from Xenopus
PT   laevis, with spasmolytic and neurotransmitter activities.
PS   Claim 1; page 5; 8pp; german.
CC   The peptide has spasmolytic and neurotransmitter functions.  Fragments
CC   comprising residues 21-400, 21-69, 351-400 and 72-348 may also be
CC   produced.  One or more of residues 1,4,5,8,9,11,20,21,22,24,25,26,33,35,
CC   36 and 42-9 may be replaced by another amino acid present in the same
CC   position in one of the other fragments, and/or contg. the exchange D-E,
CC   I-L, K-R, N-Q and/or S-T.
SQ   Sequence   400 AA;


  Query Match             7.5%;  Score 331;  DB 1;  Length 400;
  Best Local Similarity   30.1%;  Pred. No. 3.7e-11;
  Matches  114;  Conservative   17;  Mismatches  146;  Indels  102;  Gaps   18;

Qy    434 ESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAG-CQVFECDSD---------- 483
          |:|| : : |    :|             |: |    |||  :  ||  |
Db     40 EADC-RAVGCCFDSSILN----------TKWCFYNATAGPIKKLECSGDPTKRIDCGFP 87

Qy    483 ----NQC-----------------TSTAAPTAAP---TAAPTAAPTAAPTAAPSTVVPP 517
              ||                   |   | ||   ||: ||  |  ||  | |   |
Db     88 RITEKQCILRGCCFDSSISGVKWCYARTVITTPAPDTTTASTTAETTTVPT-TPETTTVP 146

Qy    518 ATPPATAAP-VPPTTAIP-TPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPE 575
           ||  |  |  | || :| ||  |  ||   ||  | :| | ||| |   |  |||  ||
Db    147 TTPETTTVPTTPETTTVPTTPETTTVPTTPETTTVPTTPETTTVPTT--PE 204

Qy    576 IPITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPT 635
            |: ||   |   |  ||   || ||  | | |:  ||  ||    | |  |  ||
Db    205 ----TTTVPTTPET---TTVPTTPETTTVPTTPETTTASTTAETTTVPTTPET--TTEPT 255

Qy    636 AAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPV- 695
            |||    |    | |||   ||      || |   |   |    | |  |   |  | |
Db    256 TTPTTDTTPP----TLPPTPETTTETTTETTTETTTETTTETTTETTTAPPPECAADRVD 311

Qy    695 ------------------------------TTTSAPATTPEDDDIDPPLPNDPINPCV-- 723
                                        |   |||   |   :|| :  |   | :
Db    312 CGYSGITQADCEGKGCIFDSTIPETKWCFYTEAEAPARKAE-CTVDPSVRTDCGYPGITD 370

Qy    723 EECNVLPWAHADC--DKYW 739
          :||      : :|  |  |
Db    371 KECREKGCCYDECIPDVIW 389


RESULT   8
P60570
ID   P60570 standard; Protein; 844 AA.
AC   P60570;
DT   24-AUG-1991  (first entry)
DE   Sequence of the Falciparum Interspersed Repeat Antigen
DE   (FIRA).
KW   Malaria vaccine; antigen; epitope.
OS   Plasmodium falciparum.
PN   WO8601802-A.
PD   27-MAR-1986.
PF   11-SEP-1985; 006960.
PR   11-SEP-1984; AU-007067.
PR   11-SEP-1984; AU-007066.
PR   10-SEP-1985; AU-047326.
PA   (HALL-) HALL INST MED RES.
PI   Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
DR   WPI; 86-094065/14.
```

```
DR   N-PSDB; N60473.
PT   DNA coding for Plasmodium falciparum antigens - expressing
PT   poly:peptide(s) having antigenicity of RESA or FIRA antigens of P
PT   falciparum
PS   Disclosure; Fig 7; 55pp; English.
CC   The inventors claim a novel DNA molecule which comprises a
CC   nucleotide sequence corresp. to all or a portion of the base
CC   sequence coding RESA (N60472) or FIRA (N60473).  RESA and FIRA have
CC   antigenicity suitable for providing protective immunity against
CC   Plasmodium falciparum malarial infections.
SQ   Sequence   844 AA;

  Query Match             7.4%;  Score 327;  DB 1;  Length 844;
  Best Local Similarity   22.5%;  Pred. No. 1.3e-10;
  Matches  183;  Conservative   30;  Mismatches  242;  Indels  360;  Gaps   34;

Qy      96 PAET--TQAPATTQAPTTTQAPTTTTQAPTT-----TTQAPTT-----TTQAPTTTQAPT 143
           | ||  || | | : | ||| |  ||| | |     | | | |     | | | | | |
Db     164 PQETVETQEPVTIEEPVTTQEP-VTTQEPITVQEPVTVQEPVTVQEPVTVQEPVTVQEPV 222

Qy     144 TTQAPTTTQAPTT-----TTQAPTT-----TTQAPTT-----TTQAPTTTQAPTTTQAPT 188
           | | | | | | |     | | | |     | | | |     ||| | ||| | | : |
Db     223 TVQEPVTVQEPVTSQEPVTPQEPVTPQEPVTPQEPVTPQEPVTTQEPVTTQEPVTIEEPV 282

Qy     189 TTQAPTTITQAATTPAATT---PAAT-TPAATTPAATTPAATTPGVPAPTSAPVWPPICE 244
           ||| | || :  ||    |   |  | |  |  |    ||    |   |  |  ||
Db     283 TTQEPVTIEEPVTTQEPVTIEEPVTTQEPVTTQEPVTTQEPVTTQEPV------ 337

Qy     245 LLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECD 304
                                                 | |:   |
Db     337 --------------------------------TVEEHIDE-------------------- 345

Qy     305 GEISPAPPVTEGNEDEDIDIGDLL----DNGCPANFEIDWL-LPHGNRCDKYYQCVHGNL 359
                  :|:| ::| :  |     :     :  ||    ||: ||       :
Db     345 ---------KKGSEGDNISLSSLSEETEEKSHTKKKKSSWLKFGRGNKNDK------KSK 389

Qy     360 VERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPN 419.
           |::                        |  || : : ||
Db     390 NEKK-----------------------PSLESVKQNADEQ-----------------K 407

Qy     420 GCPADFSIDHLLPHESDCGQYLQCVHGQ---TIARPCPGNLHFSPATQSCESPVTAGCQV 476
             | |  |                  |: |   ||  |          |||  | | |
Db     408 EQPTDSQIS---------------VNAQDSVTIQEP--------TATQ--EPPTT----- 438

Qy     477 FECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPAT--PPATA-APVPPTTAI 533
                : |:|  ||         ||   |   |  | |   | ||  ||      |
Db     438 ------QELTATQEPTTTQETVTEQEPTT--TQETVTAQEPITTQEPVTAQEPVTTQELI 489

Qy     534 PTPAPTAAPTAAPTTAA---------------------------------------------- 551
           |  |:     |    |
Db     490 ATQEPSTTQEHADEKKASEGDNISLSRLSEETEEKSHTKKKKSSWLKFGRGNKNDKKSKN 549

Qy     551 ----------------PESPT----------TVTVPPTAAPTAAPTTAVPEIPITVTSAP 584
                           | ||          :||:    |    |||      :| ·|  |
Db     550 EKKPSLESVKQNADEQKEQPTDSQISVNAQDSVTIQEPTATQEPPTTQ----ELTATQEP 605

Qy     585 TAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTA-APTT-----AAPAPNTT--------- 630
           |          ||    |   : | |   | ||  | |      |   |:||
Db     606 TTTQETVTEQEPTTTQETVTAQEPITTQEPVTAQEPVTTQELIATQEPSTTQEHADEKKA 665

Qy     630 ----------------------------------------------VTVPPTA---- 637
                                                          :|   |
Db     666 SEGDNISLSRLSEETEEKSHTKKKKSSWLKFGRGNKNDKKSKNEKKPSLESVKQNADEQK 725

Qy     637 -APTTAAPAPN----------TTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTV-A 684
            || :  : |           |    |||    ||   | ||          ||    || |
Db     726 EQPTDSQISVNAQDSVTIQEPTATQEPPTTQELTATQEPTTTQETVTEQEPTTTQETVTA 785

Qy     685 HAPNTTAAPVT-----------TTSAPATTPEDDD 708
           |  ||  |||          |  |:|| |  |
Db     786 QEPITTQEPVTAQEPVTTQELIATQEPSTTQEHAD 820

RESULT   9
P50073
ID   P50073 standard; Protein; 907 AA.
AC   P50073;
DT   17-OCT-1991  (first entry)
DE   Epstein-Barr virus (EBV) outer surface protein.
KW   Epstein-Barr virus; antigen; vaccine.
OS   Epstein-Barr virus.
PN   EP-151079-A.
PD   07-AUG-1985.
PF   28-JAN-1985; 400141.
PR   30-JAN-1984; US-575352.
PR   23-JUL-1984; US-633558.
PA   (UYCH-) UNIV OF CHICAGO.
PI   Kleff E, Tanner J, Hummel M, Belsel C;
DR   WPI; 85-191978/32.
DR   N-PSDB; N50114.
PT   New fragment of Epstein-Barr Virus DNA - useful in vector to
PT   express polypeptide for use in prepn. of vaccine against the
PT   virus and for use in diagnosis.
PS   Claim 2; Page 23-25; 26pp; English.
CC   The sequence encodes an outer surface viral protein of EBV, used
CC   to generate antibodies reacting with the surface proteins of
CC   EBV-infected cells, and in the preparation of a vaccine against EBV.
SQ   Sequence   907 AA;

  Query Match             7.4%;  Score 325;  DB 1;  Length 907;
  Best Local Similarity   23.2%;  Pred. No. 1.7e-10;
  Matches  161;  Conservative   59;  Mismatches  192;  Indels  282;  Gaps   29;

Qy     100 TQAPATTQAPTTTQAPTTTTQAPTTTTQA---PTTTTQAPTTTQAPTTTQAPTTTQAPTT 156
           | |  ||     ::|| :|| :||  |     | |||  |::|  ||   || :|  ||
Db     410 TNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPAST-GPTV 468

Qy     157 TTQAPTTTTQAPTTTTQAPTT----------TQAP--TTTQAPTTTQAP--TTITQAAT 201
           :|   |: | | ||:  :| |          ::||  |:: :| ||  |  |: | | |
Db     469 STADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTSSTSPVTTPTPNATSPTPAVT 528

Qy     202 TPAATTPAAT--TPAATTPAATTPAATTP--GVPAPTSAPVWPPICELLPNGCPADFDIH 257
           ||   || ||  ||| |||   || ||:|  |  :|||   |      ||
Db     529 TP---TPNATSPTPAVTTP---TPNATSPTLGKTSPTSAVTTP-----TPN--------- 569

Qy     258 LLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGN 317
                                                 ||      |: ||   ||
Db     569 ------------------------------------ATSPTL----GKTSPTSAVT--- 585

Qy     318 EDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQC 377
                                   |  |
Db     585 ------------------------TPTPN------------------------------ 590

Qy     378 DHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDC 437
             | |                |   ||  :   | ||
Db     590 -------ATSP-------------TLGKTSPTSAVT-TPTPNAT--------------- 613

Qy     438 GQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTA 497
               | |:    |                        : :: |      :||   |:
Db     613 --------GPTVGETSP----------------------QANATNHTLGGTSPTPVVTS 641

Qy     498 APTAAPTAA-----------------PTAAPSTVVPPATPPATAAPVPPTTAIPT---- 536
             |  | :|                 |:: | |: |  :  :|:     |:| ||
Db     642 QPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSDNSTSHMPLLTSAHPTGGEN 701

Qy     536 ---PAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVP-EIPITVTSAP-----TA 586
                | :  |   :|::|| |   |    : |  : |:  | |: :|  : |
Db     702 ITQVTPASISTHHVSTSSPE-PRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNATSPQ 760

Qy     587 APTAAPTAAPTAAPTTAVPEIPT----------TVTSPPT---AAPTTAAPAPNTTVTV 632
           ||:   || ||   |       |          | | | |      ||  |  | | :
Db     761 APSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRPRYNATTYL 820

Qy     633 PPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAA 692
```

```
          ||:   |::   |  | |:||    |||      || ||||: |  :  ::      : |
Db    821 PPS---TSSKLRPRWTFTSPPV---TTA----QATVPVPPTSQPRFSNLSMLVLQWASLA 870

Qy    693 PVT---------------TTSAPATTPEDDDID 710
           :|               :||   |||  || :
Db    871 VLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAE 904


RESULT  10
R80144
ID   R80144 standard; Protein; 907 AA.
AC   R80144;
DT   18-JAN-1996  (first entry)
DE   EBV gp350/220.
KW   EBV; gp350; gp220; gp350/gp220; non-splicing variant; vaccine.
OS   Epstein-Barr virus.
FH   Key             Location/Qualifiers
FT   peptide         1. .18
FT                   /label= Sig_peptide
FT   region          861. .881
FT                   /note= "transmembrane region"
PN   WO9528488-A1.
PD   26-OCT-1995.
PF   13-APR-1995; U04611.
PR   18-APR-1994; US-229291.
PA   (AVIR-) AVIRON.
PI   Jackman WT,  Spaete R;
DR   WPI; 95-373802/48.
DR   N-PSDB; T04821.
PT   New DNA encoding a homogeneous gp350 protein - can be used for
PT   preventing and treating Epstein-Barr virus-related diseases or
PT   conditions
PS   Disclosure; Fig.1; 61pp; English.
CC   The donor and acceptor splice sites of the EBV gene encoding gp350/
CC   220 are mutated by replacement of native nucleotides by non-native
CC   nucleotides, without altering the encoded amino acid sequence,
CC   resulting in elimination of gp220 prodn.  Recombinant homogeneous
CC   gp350, useful in vaccines, is expressed in mammalian or insect cell
CC   hosts.
SQ   Sequence   907 AA;


  Query Match             7.4%;  Score 325;  DB 1;  Length 907;
  Best Local Similarity   23.2%;  Pred. No. 1.7e-10;
  Matches  161;  Conservative   59;  Mismatches  192;  Indels  282;  Gaps   29;

Qy    100 TQAPATTQAPTTTQAPTTTTQAPTTTTQA---PTTTTQAPTTTQAPTTTQAPTTTQAPTT 156
          | |  ||     ::|| :|| :||  |     | |||  |::|  ||   || :|  ||
Db    410 TNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPAST-GPTV 468

Qy    157 TTQAPTTTTQAPTTTTQAPTT----------TQAP--TTTQAPTTTQAP--TTITQAAT 201
          :|   |: | | ||:  :| |           ::||  |:: :| ||  |  |: | | |
Db    469 STADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTSSTSPVTTPTPNATSPTPAVT 528

Qy    202 TPAATTPAAT--TPAATTPAATTPAATTP--GVPAPTSAPVWPPICELLPNGCPADFDIH 257
          ||   || ||  ||| |||   || ||:|  |  :|||   |      ||
Db    529 TP---TPNATSPTPAVTTP---TPNATSPTLGKTSPTSAVTTP-----TPN--------- 569

Qy    258 LLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGN 317
                                              ||      |: ||   ||
Db    569 ------------------------------------ATSPTL----GKTSPTSAVT--- 585

Qy    318 EDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQC 377
                                  | |
Db    585 ------------------------TPTPN------------------------------ 590

Qy    378 DHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDC 437
                | |               |   ||  :   | ||
Db    590 -------ATSP------------TLGKTSPTSAVT-TPTPNAT---------------- 613

Qy    438 GQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTA 497
                | |:   |                        : :: |      :||   |:
Db    613 --------GPTVGETSP----------------------QANATNHTLGGTSPTPVVTS 641

Qy    498 APTAAPTAA-----------------PTAAPSTVVPPATPPATAAPVPPTTAIPT---- 536
          |  | :|                   |:: | |: |  :  :|:     |:| ||
Db    642 QPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSDNSTSHMPLLTSAHPTGGEN 701

Qy    536 ---PAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVP-EIPITVTSAP-----TA 586
              | :  |   :|::|| |   |    :  | : |:  | |: :|  : |
Db    702 ITQVTPASISTHHVSTSSPE-PRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNATSPQ 760

Qy    587 APTAAPTAAPTAAPTTAVPEIPT----------TVTSPPT---AAPTTAAPAPNTTVTV 632
          ||:   || ||   |       |           | | | |      ||  |  | |  :
Db    761 APSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRPRYNATTYL 820

Qy    633 PPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAA 692
          ||:   |::   |  | |:||    |||      || ||||: |  :  ::      : |
Db    821 PPS---TSSKLRPRWTFTSPPV---TTA----QATVPVPPTSQPRFSNLSMLVLQWASLA 870

Qy    693 PVT---------------TTSAPATTPEDDDID 710
           :|               :||   |||  || :
Db    871 VLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAE 904


RESULT  11
R07671
ID   R07671 standard; protein; 278 AA.
AC   R07671;
DT   22-FEB-1991  (first entry)
DE   Intestinal mucin deduced from clone SMUC 41.
KW   Epithelial cancer; cystic fibrosis; ulcerative colitis;
KW   glycosyl transferase; familial polyposis coli;
KW   glycosylation patterns.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   region          6. .28
FT                   /label= repeat 1
FT   region          29. .52
FT                   /label= repeat 2
FT   region          53. .75
FT                   /label= repeat 3
FT   region          76. .98
FT                   /label= repeat 4
FT   region          99. .121
FT                   /label= repeat 5
FT   modified_site   265. .267
FT                   /label= N-glycosylation site
PN   WO9012892-A.
PD   01-NOV-1990.
PF   25-JUL-1989; U03206.
PR   14-APR-1989; US-338710.
PA   (REGC) University of California.
PI   Kim YS; Gum JR
DR   WPI; 90-348495/46.
DR   N-PSDB; Q06484.
PT   Human intestinal mucin DNA, polypeptides and antibodies - used in
PT   diagnosis and treatment of e.g. cancers, cystic fibrosis and
PT   colitis.
PS   Disclosure; Fig 4B; 45pp; English.
CC   The sequence was deduced from a clone isolated from a cDNA library
CC   using antibody probes. The DNA sequence allows determination of the
CC   primary amino acid structure, difficult to characterise
CC   biochemically because the protein backbone is so heavily glycosyl-
CC   ated and conditions required to remove the carbohydrate tend to
CC   result in breakage of the protein.  Prods. of the DNA can be used
CC   to produce Abs useful for detecting hIM and determining glycosyl-
CC   ation patterns or glycosyltransferase activity.  They can be used
CC   for the early detection and differential diagnosis of cancers esp.
CC   epithelial cancer, and also in diseases with altered intestinal
CC   mucin prodn. such as cystic fibrosis and colitis.
CC   See also R07670-76.
SQ   Sequence   278 AA;


  Query Match             7.1%;  Score 313.5;  DB 1;  Length 278;
```

```
  Best Local Similarity   38.3%;  Pred. No. 2.2e-10;
  Matches   82;  Conservative   19;  Mismatches   88;  Indels   25;  Gaps    8;

Qy    502 APTAAPTAAPSTVVPPATPPATAAPVP---PTTAIPTPAPTAAPTAAPTTAAPESPTTVT 558
          |||  | :  :|| |  ||  |  |      ||   || ||   |  |||      || |
Db      6 APTPTPISTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTVL--ITTTTT 63

Qy    559 VPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAA 618
          : ||  ||:  :| |   ||| |:  || ||   |  ||  |      | || |  ||
Db     64 MTPTPTPTSTKSTTV--TPITTTTTVTATPTPTGTQTPTMIP------ISTTTTVTPTPT 115

Qy    619 PTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTV--TVPPTAAP 676
          ||| :  | |  :  | |  ||: | | :  : | |:  |::    :||:  |:||
Db    116 PTTGSTGPPTHTSTAPIAELTTSNPPPES--STPQTSRSTSSPLTESTTLLSTLPPAIEM 173

Qy    677 TA-APPTVAHAPNTTAA-------PVTTTSAPAT 702
          |: |||:   || ||         | ||||  | |
Db    174 TSTAPPSTPTAPTTTTGAHTLSPPPSTTTSPPGT 207


RESULT  12
W64359
ID   W64359 standard; Protein; 267 AA.
AC   W64359;
DT   09-NOV-1998  (first entry)
DE   Mycobacterium tuberculosis antigen TbH-29.
KW   Tuberculosis; infection; diagnosis; antigen; TbH-29.
OS   Mycobacterium tuberculosis strain H37Rv.
PN   WO9816645-A2.
PD   23-APR-1998.
PF   07-OCT-1997; U18214.
PR   13-MAR-1997; US-818111.
PR   11-OCT-1996; US-729622.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI   Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR   WPI; 98-251292/22.
DR   N-PSDB; V44403.
PT   New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT   to develop products for the detection of M. tuberculosis infection
PT   and diagnosis of tuberculosis
PS   Example 3; Page 157-158; 250pp; English.
CC   This polypeptide comprises an antigenic portion of Mycobacterium
CC   tuberculosis antigen TbH-29.  A DNA molecule (see V44403) encoding
CC   the polypeptide was isolated from a M. tuberculosis strain H37Rv
CC   genomic library.  The invention relates to compositions and methods
CC   for diagnosing tuberculosis.  It provides polypeptides (see
CC   W64291-W64379) comprising an antigenic portion of a soluble M.
CC   tuberculosis antigen, or an immunogenic portion of an M.
CC   tuberculosis antigen, as well as DNA sequences encoding such
CC   polypeptides, recombinant expression vectors and transformed or
CC   transfected host cells.  Also claimed are methods and diagnostic
CC   kits for detecting M. tuberculosis infection in a patient using
CC   these polypeptides, antibodies or oligonucleotide probes and
CC   primers, for the diagnosis of tuberculosis.
SQ   Sequence   267 AA;


  Query Match             7.0%;  Score 309.5;  DB 1;  Length 267;
  Best Local Similarity   34.8%;  Pred. No. 3.5e-10;
  Matches   86;  Conservative   11;  Mismatches  107;  Indels   43;  Gaps    8;

Qy    487 STAAPTAAPTAAPTAAPTAAPTAAPSTVVP----PATPPATAAPVPP------------ 530
          :: |   | |  ||||  : |        |    |  |    |||||
Db     12 ASLAVAVAVTIRPTAA--SKPVEGHQNAQPGKFMPLLPTQQQAPVPPPPPDDPTAGFQGG 69

Qy    530 -----TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTA------VPEIPI 578
                :| |  :      | : |||:|    | |   |   |          | :|
Db     70 TIPAVQNVVPRPGTSPGVGGTPASPAPEAPAVPGVVPAPVPIPVPIIIPPPFPGWQPGMPT 129

Qy    579 TVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTT-----VTSPPTAAPTTAAPAPNTTVTVP 633
            |: ||   | : |  ||  ||| |   |||     ||:|||  |||    | |||  |
Db    130 IPTAPPTTPVTTSATTPPTTPPTTPVTTPPTTPPTTPVTTPPTTPPTTPVTTPPTTV-AP 188

Qy    634 PTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAP 693
          | |||| ||    |  || | || || |   |||   ||  ||  |         | ||
Db    189 TTVAPTTVAP----TTVAPTTVAPATATP---TTVAPQPTQQPTQQPTQQMPTQQQTVAP 241

Qy    694 VTTTSAP 700
            |   ||
Db    242 QTVAPAP 248


RESULT  13
W81726
ID   W81726 standard; Protein; 267 AA.
AC   W81726;
DT   27-JAN-1999  (first entry)
DE   M. tuberculosis immunogenic polypeptide TbH-29.
KW   Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW   vaccine; pharmaceutical; infection; diagnosis.
OS   Mycobacterium tuberculosis.
PN   WO9816646-A2.
PD   23-APR-1998.
PF   07-OCT-1997; U18293.
PR   13-MAR-1997; US-818112.
PR   11-OCT-1996; US-730510.
PA   (CORI-) CORIXA CORP.
PI   Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI   Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
DR   WPI; 98-261042/23.
DR   N-PSDB; V64512.
PT   Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
PT   to develop products for the detection of M. tuberculosis infection
PT   and for diagnosis, treatment and prevention of tuberculosis
PS   Example 3b; Page 149-150; 230pp; English.
CC   This sequence represents an immunogenic portion of a soluble
CC   Mycobacterium tuberculosis (MT) antigen which can be used in a method
CC   for inducing protective immunity against tuberculosis (TB). This
CC   sequence can be formulated into vaccines and/or pharmaceutical
CC   compositions for immunising against M. tuberculosis infection or may
CC   be used for the diagnosis of tuberculosis.
SQ   Sequence   267 AA;


  Query Match             7.0%;  Score 309.5;  DB 1;  Length 267;
  Best Local Similarity   34.8%;  Pred. No. 3.5e-10;
  Matches   86;  Conservative   11;  Mismatches  107;  Indels   43;  Gaps    8;

Qy    487 STAAPTAAPTAAPTAAPTAAPTAAPSTVVP----PATPPATAAPVPP------------ 530
          :: |   | |  ||||  : |        |    |  |    |||||
Db     12 ASLAVAVAVTIRPTAA--SKPVEGHQNAQPGKFMPLLPTQQQAPVPPPPPDDPTAGFQGG 69

Qy    530 -----TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTA------VPEIPI 578
                :| |  :      | : |||:|    | |   |   |          | :|
Db     70 TIPAVQNVVPRPGTSPGVGGTPASPAPEAPAVPGVVPAPVPIPVPIIIPPPFPGWQPGMPT 129

Qy    579 TVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTT-----VTSPPTAAPTTAAPAPNTTVTVP 633
            |: ||   | : |  ||  ||| |   |||     ||:|||  |||    | |||  |
Db    130 IPTAPPTTPVTTSATTPPTTPPTTPVTTPPTTPPTTPVTTPPTTPPTTPVTTPPTTV-AP 188

Qy    634 PTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAP 693
          | |||| ||    |  || | || || |   |||   ||  ||  |         | ||
Db    189 TTVAPTTVAP----TTVAPTTVAPATATP---TTVAPQPTQQPTQQPTQQMPTQQQTVAP 241

Qy    694 VTTTSAP 700
            |   ||
Db    242 QTVAPAP 248


RESULT  14
R92803
ID   R92803 standard; Protein; 451 AA.
AC   R92803;
DT   24-MAY-1996  (first entry)
DE   Hepatitis A virus receptor.
```

```
KW   Hepatitis A virus; receptor; HAV; diagnosis; therapy; vaccine;
KW   transgenic animal.
OS   Cercopithecus aethiops.
PN   WO9604376-A1.
PD   15-FEB-1996.
PF   04-AUG-1995; U09941.
PR   05-AUG-1994; US-287001.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Feinstone SM,  Kaplan G;
DR   WPI; 96-129392/13.
DR   N-PSDB; T17864.
PT   DNA encoding Hepatitis A virus receptor and its fragments - useful
PT   for the detection, diagnosis, prevention and treatment of HAV
PT   infection
PS   Claim 1; Page 39-41; 52pp; English.
CC   A hepatitis A virus (HAV) receptor (R92803) was identified as the
CC   product of a nucleic acid sequence isolated from HAV-infected
CC   primary African green monkey kidney cells.  Recombinant receptor
CC   can be produced in large amounts by expression of the nucleic acid
CC   in prokaryotic or eukaryotic host cells or in transgenic animals.
CC   It can be used in the detection, prevention and treatment of HAV
CC   infection, in methods of purifying or removing HAV from samples,
CC   and methods of determining the anti-HAV activity of cpds.
CC   Expression in transgenic animals provides a method for testing
CC   vaccine efficacy.
SQ   Sequence   451 AA;
```

```
  Query Match             7.0%;  Score 307.5;  DB 1;  Length 451;
  Best Local Similarity   41.6%;  Pred. No. 7.5e-10;
  Matches   87;  Conservative    4;  Mismatches   53;  Indels   65;  Gaps   10;

Qy      94 GPPAET----------TQAPATTQAPTTTQAPTTTQAPTTTQAPTT----TTQAPTTT 139
           |||  |          |  | ||  ||||  |||||  |||||   ||    ||  ||||
Db     125 GPPRVTTPIVRTVRTSTTVPTTTTLPTTTTLPTTTT-LPTTTTLPMTTTLPMTTTHPTTT 183

Qy     140 QAPTTTQAPTTTQAPTT---TTQAPTTTTQAPTTTT-----QAPTTTQAPTTTQAPTTTQ 191
             ||||  ||||  |||   ||  ||| |  |||||       | ||  ||||  ||||
Db     184 TVPTTTTLPTTTTLPTTLPMTTTLPTTRT-LPTTTTLPTTMTLPMTTTLPTTTTLPTTTT 242

Qy     192 APTTITQAATTPAATT-------PAATT-PAATTPAATT----------------PAAT 226
            |||     | |  ||       |  || |  ||   ||                | ||
Db     243 LPTTTLPTMTLPTTTTLPTMMTLPTTTTLPTTTTLPTTTMVSTFVPPTPLPMQNHEPVAT 302

Qy     227 TPGVP----------------APTSAPVW 239
           :|  |                |||:|::
Db     303 SPSSPQPAETHPVTLLGATRTQPTSSPLY 331
```

```
RESULT  15
R12535
ID   R12535 standard; Protein; 278 AA.
AC   R12535;
DT   02-SEP-1991  (first entry)
DE   SMUC-41 intestinal mucin.
KW   SMUC; intestinal mucin; diagnosis; cancer.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   region          7. .29
FT                   /label= repeat_1
FT   region          30. .52
FT                   /label= repeat_2
FT   cross_links     53. .75
FT                   /label= repeat_3
FT   region          76. .98
FT                   /label= repeat_4
FT   region          99. .121
FT                   /label= repeat_5
FT   modified_site   265. .267
FT                   /label= glycosylation_site
PN   WO9108217-A.
PD   13-JUN-1991.
PF   04-DEC-1990; U07087.
PR   05-DEC-1989; US-447140.
PA   (REGC ) UNIV OF CALIFORNIA.
PI   Kim YS, Gum JR;
DR   WPI; 91-193147/26.
DR   N-PSDB; Q12218.
PT   Nucleic acid encoding human intestinal mucin peptide(s) - used
PT   for producing prods. for diagnosis and treatment of disorders
PT   such as cancer
PS   Disclosure; Fig 2(B); 57pp; English.
CC   The sequence contains repeats having a threonine and proline-rich
CC   consensus.
CC   Antibodies produced against this mucin peptide can be used for
CC   cancer therapy by coupling to toxins and can also be used in treatment,
CC   partic. by blocking or interrupting some functional activity of
CC   intestinal mucin protein which contributes to the neoplastic phenotype.
CC   A vaccination for inducing a cellular immune response can be prepd.  The
CC   prods. are also useful as substrates for assay of glycosyltransferase
CC   activity and other glycosylation enzymes.
CC   See also Q12217-21 and Q12247-51.
SQ   Sequence   278 AA;
```

```
  Query Match             6.9%;  Score 303;  DB 1;  Length 278;
  Best Local Similarity   33.8%;  Pred. No. 8.1e-10;
  Matches   79;  Conservative   26;  Mismatches  109;  Indels   20;  Gaps    8;

Qy     486 TSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVP---PTTAIPTPAPTAAP 542
           |:|  ||  ||   |  ||  |    :|| |  ||  |  |      ||   || ||  |
Db      15 TTTVTPTPTPTG--TQTPTTTPITTTTTVTPTPTPTGTQTPTTVLITTTTTMTPTPT--P 70

Qy     543 TAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTT 602
           |:  :|      || ||  |  ||   |  :  |  | |  ||  ||   |  ||   |
Db      71 TSTKSTTVTPITTTTTVTATRTPTGTQTPTMIPISTTTTVTPTPTPTTGSTGPPTHTSTA 130

Qy     603 AVPEIPTTVTSPPTAAP----TTAAPAPNTTV---TVPPTAAPTTAAPAPNTTVTAPPTA 655
            : |: |:   | :: |    :|::|   :|    |:|| |   |:  ||:|      |:
Db     131 PIAELTTSNPPPESSTPQTSRSTSSPLTESTTLLSTLPP-AIEMTSTAAPSTPTAPTTTS 189

Qy     656 APTTAAPAPNTTVTVP--PTAAPT---AAPPTVAHAPNTTAAPVTTTSAPATTP 704
           | :| |:|| : |  ||   |   :: || :    || :  | |  | :||
Db     190 GGHTLSPPPSTTTSPPGTPTRGTTTGSSSAPTPSTVQTTTTSAWTPTPTPLSTP 243
```

```
Search completed: July 18, 1999, 00:46:15
Job time: 268 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 07:57:21 ; Search time 21.52 Seconds
(without alignments)
370.064 Million cell updates/sec

Title: US-09-294-663-4
Perfect score: 4395
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 807

Scoring table: BLOSUM62

Searched: 106577 seqs, 9868381 residues

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/PCTUS9_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	325	7.4	907	3	PCT-US95-04611A-19	Sequence 19, Appl
2	308.5	7.0	451	1	US-08-287-001A-2	Sequence 2, Appli
3	308.5	7.0	451	3	PCT-US95-09941-2	Sequence 2, Appli
4	283.5	6.5	2476	2	US-08-276-967-2	Sequence 2, Appli
5	276.5	6.3	214	1	US-08-217-327-4	Sequence 4, Appli
6	275.5	6.3	2035	2	US-08-479-537A-2	Sequence 2, Appli
7	275.5	6.3	1867	2	US-08-479-537A-5	Sequence 5, Appli
8	247	5.6	1537	1	US-08-325-267A-2	Sequence 2, Appli
9	239	5.4	486	1	US-08-450-360-2	Sequence 2, Appli
10	236.5	5.4	895	1	US-08-123-161A-8	Sequence 8, Appli
11	236.5	5.4	895	1	US-08-483-278-8	Sequence 8, Appli
12	232.5	5.3	2035	1	US-08-046-585-5	Sequence 5, Appli
13	232.5	5.3	2035	1	US-08-393-703-5	Sequence 5, Appli
14	232.5	5.3	2035	3	PCT-US93-11721-5	Sequence 5, Appli
15	224	5.1	357	1	US-08-078-683A-8	Sequence 8, Appli
16	219.5	5.0	447	1	US-08-450-360-4	Sequence 4, Appli
17	219.5	5.0	629	3	PCT-US95-07721-6	Sequence 6, Appli
18	214.5	4.9	722	1	US-08-347-718B-1	Sequence 1, Appli
19	214.5	4.9	742	1	US-08-347-718B-2	Sequence 2, Appli
20	214.5	4.9	745	2	US-08-445-050-2	Sequence 2, Appli
21	214.5	4.9	722	2	US-08-445-050-3	Sequence 3, Appli
22	214.5	4.9	722	2	US-08-445-050-7	Sequence 7, Appli
23	214.5	4.9	722	2	US-08-482-262-1	Sequence 1, Appli
24	214.5	4.9	742	2	US-08-482-262-2	Sequence 2, Appli
25	214.5	4.9	745	2	US-08-204-691-2	Sequence 2, Appli
26	214.5	4.9	722	2	US-08-204-691-3	Sequence 3, Appli
27	214.5	4.9	722	2	US-08-204-691-7	Sequence 7, Appli
28	207.5	4.7	862	1	US-08-325-267A-4	Sequence 4, Appli
29	207	4.7	806	1	US-08-270-076A-11	Sequence 11, Appl
30	205	4.7	985	3	PCT-US96-03916-6	Sequence 6, Appli
31	205	4.7	985	3	PCT-US96-03916-66	Sequence 66, Appl
32	204.5	4.7	1848	3	PCT-US95-10661A-6	Sequence 6, Appli
33	202.5	4.6	234	2	US-08-642-255-51	Sequence 51, Appl
34	202.5	4.6	561	2	US-08-642-255-52	Sequence 52, Appl
35	198	4.5	1481	2	US-08-616-844-40	Sequence 40, Appl
36	198	4.5	1481	2	US-08-599-654-40	Sequence 40, Appl
37	197	4.5	1706	2	US-08-459-568-2	Sequence 2, Appli
38	197	4.5	1706	2	US-08-399-411-2	Sequence 2, Appli
39	196	4.5	2441	2	US-08-194-468-2	Sequence 2, Appli
40	196	4.5	1719	2	US-08-459-568-4	Sequence 4, Appli
41	196	4.5	1719	2	US-08-399-411-4	Sequence 4, Appli
42	195.5	4.4	921	1	US-08-188-582-2	Sequence 2, Appli
43	195.5	4.4	921	1	US-08-646-715-2	Sequence 2, Appli
44	192.5	4.4	412	2	US-08-477-254A-4	Sequence 4, Appli
45	192.5	4.4	412	2	US-08-472-576B-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT   1
PCT-US95-04611A-19
; Sequence 19, Application PC/TUS9504611A
;  GENERAL INFORMATION:
;    APPLICANT:  Spaete, Richard and Jackman, Winthrop, T.
;    TITLE OF INVENTION:  Non Splicing Variants of gp350/220
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooley Godward Castro Huddleson & Tatum
;      STREET:  5 Palo Alto Square
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04611A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/229,291
;      FILING DATE:  April 18, 1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  AVIR-003/00US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-843-5163
;      TELEFAX:  415-857-0663
;      TELEX:  380816 CooleyPA
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  907 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-04611A-19


  Query Match             7.4%;  Score 325;  DB 3;  Length 907;
  Best Local Similarity   23.2%;  Pred. No. 2.4e-14;
  Matches  161;  Conservative   59;  Mismatches  192;  Indels  282;  Gaps   29;

Qy    100 TQAPATTQAPTTTQAPTTTTQAPTTTTQA---PTTTTQAPTTTQAPTTTQAPTTTQAPTT 156
          | |  ||      ::|| :|| :||  |      | |||  |::|  ||   || :|  ||
Db    410 TNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPAST-GPTV 468

Qy    157 TTQAPTTTTQAPTTTTQAPTT-----------TQAP--TTTQAPTTTQAP--TTITQAAT 201
          :|   |: | | ||:  :| |           ::||  |:: :| ||  |  |: | | |
Db    469 STADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTSSTSPVTTPTPNATSPTPAVT 528

Qy    202 TPAATTPAAT--TPAATTPAATTPAATTP--GVPAPTSAPVWPPICELLPNGCPADFDIH 257
          ||   || ||  ||| |||   || ||:|  |  :|||   |      ||
Db    529 TP---TPNATSPTPAVTTP---TPNATSPTLGKTSPTSAVTTP-----TPN--------- 569

Qy    258 LLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGN 317
                                                 ||      |: ||   ||
```

```
Db    569 ------------------------------------ATSPTL----GKTSPTSAVT--- 585

Qy    318 EDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQC 377

Db    585 ------------------------TPTPN------------------------------ 590

Qy    378 DHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDC 437

Db    590 -------ATSP-------------TLGKTSPTSAVT-TPTPNAT---------------- 613

Qy    438 GQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTA 497

Db    613 --------GPTVGETSP----------------------QANATNHTLGGTSPTPVVTS 641

Qy    498 APTAAPTAA-----------------PTAAPSTVVPPATPPATAAPVPPTTAIPT---- 536

Db    642 QPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSDNSTSHMPLLTSAHPTGGEN 701

Qy    536 ---PAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVP-EIPITVTSAP-----TA 586

Db    702 ITQVTPASISTHHVSTSSPE-PRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNATSPQ 760

Qy    587 APTAAPTAAPTAAPTTAVPEIPT-----------TVTSPPT---AAPTTAAPAPNTTVTV 632

Db    761 APSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRPRYNATTYL 820

Qy    633 PPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAA 692

Db    821 PPS---TSSKLRPRWTFTSPPV---TTA----QATVPVPPTSQPRFSNLSMLVLQWASLA 870

Qy    693 PVT----------------TTSAPATTPEDDDID 710

Db    871 VLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAE 904
```

```
RESULT   2
US-08-287-001A-2
; Sequence 2, Application US/08287001A
; Patent No. 5622861
;  GENERAL INFORMATION:
;    APPLICANT:  KAPLAN, GERARDO
;    APPLICANT:  FEINSTONE, STEPHEN M.
;    TITLE OF INVENTION:  HEPATITIS A VIRUS RECEPTOR AND METHODS
;    TITLE OF INVENTION:  OF USE
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg, 127 Peachtree
;      STREET:  Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/287,001A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spratt, Gwedolyn D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  1414.621
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  451 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-287-001A-2


  Query Match             7.0%;  Score 308.5;  DB 1;  Length 451;
  Best Local Similarity   41.6%;  Pred. No. 1.4e-13;
  Matches   87;  Conservative    4;  Mismatches   53;  Indels   65;  Gaps   10;

Qy     94 GPPAET----------TQAPATTQAPTTTQAPTTTTQAPTTTTQAPTT----TTQAPTTT 139

Db    125 GPPRVTTPIVRTVRTSTTVPTTTTLPTTTTLPTTTT-LPTTTTLPMTTTLPMTTTLPTTT 183

Qy    140 QAPTTTQAPTTTQAPTT---TTQAPTTTTQAPTTTT-----QAPTTTQAPTTTQAPTTTQ 191

Db    184 TVPTTTTLPTTTTLPTTLPMTTTLPTTRT-LPTTTTLPTTMTLPMTTTLPTTTTLPTTTT 242

Qy    192 APTTITQAATTPAATT-------PAATT-PAATTPAATT-----------------PAAT 226

Db    243 LPTTTLPTMTLPTTTTLPTMMTLPTTTTLPTTTTLPTTTMVSTFVPPTPLPMQNHEPVAT 302

Qy    227 TPGVP----------------APTSAPVW 239

Db    303 SPSSPQPAETHPVTLLGATRTQPTSSPLY 331


RESULT   3
PCT-US95-09941-2
; Sequence 2, Application PC/TUS9509941
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  HEPATITIS A VIRUS RECEPTOR AND METHODS
;    TITLE OF INVENTION:  OF USE
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NEEDLE & ROSENBERG, P.C.
;      STREET:  Suite 1200, The Candler Bldg, 127 Peachtree
;      STREET:  Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/09941
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/287,001
;      FILING DATE:  5 AUG 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spratt, Gwedolyn D.
;      REGISTRATION NUMBER:  36,016
;      REFERENCE/DOCKET NUMBER:  1414.621
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  404/688-0770
;      TELEFAX:  404/688-9880
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  451 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-09941-2


  Query Match             7.0%;  Score 308.5;  DB 3;  Length 451;
  Best Local Similarity   41.6%;  Pred. No. 1.4e-13;
  Matches   87;  Conservative    4;  Mismatches   53;  Indels   65;  Gaps   10;
```

```
Qy      94 GPPAET----------TQAPATTQAPTTTQAPTTTTQAPTTTTQAPTT----TTQAPTTT 139
           |||  |          |  | ||  ||||  |||||  |||||   ||    ||  ||||
Db     125 GPPRVTTPIVRTVRTSTTVPTTTTLPTTTTLPTTTT-LPTTTTLPMTTTLPMTTTLPTTT 183

Qy     140 QAPTTTQAPTTTQAPTT---TTQAPTTTTQAPTTTT-----QAPTTTQAPTTTQAPTTTQ 191
            ||||  ||||  |||   ||  ||| |  |||||       | ||  ||||  ||||
Db     184 TVPTTTTLPTTTTLPTTLPMTTTLPTTRT-LPTTTTLPTTMTLPMTTTLPTTTTLPTTTT 242

Qy     192 APTTITQAATTPAATT-------PAATT-PAATTPAATT----------------PAAT 226
           |||     | |  ||        |  || |  ||   ||                | ||
Db     243 LPTTTLPTMTLPTTTTLPTMMTLPTTTTLPTTTTLPTTTMVSTFVPPTPLPMQNHEPVAT 302

Qy     227 TPGVP----------------APTSAPVW 239
           :|  |                |||:|::
Db     303 SPSSPQPAETHPVTLLGATRTQPTSSPLY 331


RESULT   4
US-08-276-967-2
; Sequence 2, Application US/08276967
; Patent No. 5851817
;  GENERAL INFORMATION:
;    APPLICANT:  Hardy, Daniel M.
;    APPLICANT:  Garbers, David L.
;    TITLE OF INVENTION:  Species-Specific Egg-Binding Proteins of
;    TITLE OF INVENTION:  Sperm
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P. O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210-4433
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,967
;      FILING DATE:  Submitted Herewith
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitchell, Barbara S.
;      REGISTRATION NUMBER:  33,928
;      REFERENCE/DOCKET NUMBER:  UTSD:418\KIT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-787-1400
;      TELEFAX:  713-789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2476 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-276-967-2


  Query Match             6.5%;  Score 283.5;  DB 2;  Length 2476;
  Best Local Similarity   20.9%;  Pred. No. 4.2e-11;
  Matches  152;  Conservative   42;  Mismatches  212;  Indels  323;  Gaps   31;

Qy      90 CTLPGPPAETTQAP-ATTQAPTT--------TQAPTTTTQAPTTTTQAPTTTTQAPT--T 138
           |  |   : :|: | | |: ||         |: |    : |   |: ||  |: ||  |
Db     310 CRGPSETSVSTEKPVAPTEKPTVPSEIYTIPTEKPMVHMEKPIVHTEKPTVPTEKPTIPT 369

Qy     139 TQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTT-TQAPTT-TQAPTT-TQAPTT 195
           ::   |: ||  : ||   : ||   : |||  : |    : ||  |: ||: |:  |
Db     370 EKSTVPTKKPTVFKEPTLPPEGPTVPAERPTTPPEGPAVPPKGPTVLTEWPTSHTEKSTV 429

Qy     196 ITQAATTPAATTPAAT-TPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPADF 254
           |:   |  :  |  |  |  |||  |
Db     430 HTEKPILPTGKSTIPTEKPMVPTKRTTTPTERT-------------------------- 463

Qy     255 DIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVT 314
           || :|
Db     463 ----TIPAEK---------------------------------------------------- 469

Qy     315 EGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFEL 374

Db     469 ------------------------------------------------------------ 469

Qy     375 QQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHE 434
                                           || |||   :|
Db     469 ---------------------------------PTVPIEKPMVP---------------- 480

Qy     435 SDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAA 494
                                      |:   |                |  ||
Db     480 --------------------------TERTTIPT--------------ERTTIPTEK 496

Qy     495 PTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTT--AIPTPAPTAAP--TAAP---T 547
           ||  ||   |  ||  |  :||      |  |: ||    |||   |     |  |   |
Db     497 PT-VPTEKLT-VPTEKP--IVP------TEKPIVPTEKHTIPTEKLTVLTERTTTPTERT 546

Qy     548 TAAPESPTTVTVPPTAAPTAAPT--TAVPEIP---ITVTSAPTAAPTAAPTAAPTAAPTT 602
           |   | ||  |  | : ||  ||  |  | ||   :|| :  |  ||   |  ||   ||
Db     547 TIPTEKPTVPTEKP-SVPTEKPTVPTEEPTIPTEKLTVPTERTTTPTKR-TTTPTIRTTT 604

Qy     603 AVPEIPTTVTSPPTAAPT--TAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPT-T 659
             | | ||  :  |  ||  |  |   ||:    |  ||     |     || |  |: |
Db     605 --PTIRTTTPTERTTTPTIRTTTPTERTTIPTKKTTVPTEKTIIPTERTIAPTTPQPSPT 662

Qy     660 AAPAPNTTVTVPPTAAPTAAP-PTVAHAP-----NTTAAPVTTTSAPATTPEDDDIDPP- 713
             |     | :| |:| |  |  |:|  |         | ||   |  : ||  :    |
Db     663 LVPTQPAAVVMPSTSATTVTPRTTIASCPPNAHFERCACPV---SCQSPTPNCELFCKPG 719

Qy     713 ----------------------------LPNDPINP-CVEECNVLPWAHADCDKY--- 739
                                        |  |  :| | | |:  | :  :|  :
Db     720 CVCDPGFLFSGSHCVNASSCDCFYNDNYYKLGTDWFSPNCTEHCHCRPSSRMECQTFKCG 779

Qy     739 --WVCDGNN 745
             ||   |
Db     780 THTVCQLKN 788


RESULT   5
US-08-217-327-4
; Sequence 4, Application US/08217327
; Patent No. 5474925
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal E
;    APPLICANT:  Barton, Kenneth A
;    TITLE OF INVENTION:  Immobilized Proteins in Cotton Fiber
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  P.O. Box 2113
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  USA
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/217,327
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/812,233
```

```
;      FILING DATE:  19-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J
;      REGISTRATION NUMBER:  27,386
;      REFERENCE/DOCKET NUMBER:  1122990831
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  214 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-217-327-4


  Query Match             6.3%;  Score 276.5;  DB 1;  Length 214;
  Best Local Similarity   39.2%;  Pred. No. 7.9e-12;
  Matches   80;  Conservative   12;  Mismatches   67;  Indels   45;  Gaps   11;

Qy     502 APTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPP 561
           ||:  ||: |:|  |||: |      ||||  | | ||| |           |   | ||
Db      25 APSNPPTSTPATPTPPASTP------PPTTQAP-PTPTATP----------PPVSTPPP 66

Qy     562 TAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTT 621
           |::|   | || |  |   |  |:: | | |   | |:|  | |       ||| | |
Db      67 TSSP--PPVTASP--PPVSTPPPSSPPPATP---PPASPPPATP----PPASPPPATPPP 115

Qy     622 AAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTT--AAPAPNTTVTVPPTAAPTAA 679
           |:| | |    || | |  | | | |   ||  : | |  |     | :| ||      |
Db     116 ASPPPAT----PPPATPPPATPPPATPPPAPLASPPATVPAISPVQTPLTSPP------A 165

Qy     680 PPTVAHAPNTTAAPVTTTSAPATT 703
           ||| | ||   ||    |  || |
Db     166 PPTEAPAPTLGAA----TPGPAGT 185


RESULT   6
US-08-479-537A-2
; Sequence 2, Application US/08479537A
; Patent No. 5861381
;  GENERAL INFORMATION:
;    APPLICANT:  CHAMBON, Pierre
;    APPLICANT:  KIENY, Marie-Paule
;    APPLICANT:  LATHE, Richard
;    APPLICANT:  HAREUVENI, Mara
;    TITLE OF INVENTION:  PHARMACEUTICAL COMPOSITION FOR THE
;    TITLE OF INVENTION:  TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,537A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 90/13101
;      FILING DATE:  23-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR91/00835
;      FILING DATE:  23-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/039,320
;      FILING DATE:  04-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/403,576
;      FILING DATE:  14-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  017753-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  128..1899
;      OTHER INFORMATION:  /note= "The amino acids spanning
;      OTHER INFORMATION:  128 to 1899 constitute a repeated region wherein the repea
;      OTHER INFORMATION:  20 amino acids, 17 of which are fixed. The number of such
;      OTHER INFORMATION:  repeats varies from 1 to 40."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  134
;      OTHER INFORMATION:  /note= "Amino acid 134 is X1 = Xaa
;      OTHER INFORMATION:  Xaa Xaa which is the codon for Pro or Ala wherein Pro = CC
;      OTHER INFORMATION:  CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  144
;      OTHER INFORMATION:  /note= "Amino acid 144 is Y = Xaa
;      OTHER INFORMATION:  which is the codon for Thr or Asn wherein Thr = ACT, ACC,
;      OTHER INFORMATION:  or ACG; and Asn = AAT or AAC."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  147
;      OTHER INFORMATION:  /note= "Amino acid 147 is X2 = Xaa
;      OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC,
;      OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  1..21
;      OTHER INFORMATION:  /note= "Amino acids 1 to 21 are a
;      OTHER INFORMATION:  21 amino acid precursor sequence."
US-08-479-537A-2


  Query Match             6.3%;  Score 275.5;  DB 2;  Length 2035;
  Best Local Similarity   23.5%;  Pred. No. 1.1e-10;
  Matches  166;  Conservative   45;  Mismatches  291;  Indels  203;  Gaps   35;

Qy      93 PGPPAETTQAPATTQAPTTTQAPTT-------TTQAPTTTTQAPTTTTQAPTTTQAPTTT 145
           ||  : |||    | || |  |  :        |  | |     :||  |   | ||
Db      66 PGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 125

Qy     146 QAPTTTQAPTT--TTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAP-----TTITQ 198
           || :| ||     | ||       :|   |   | ||     | :| ||     |:
Db     126 PAPGST-APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGST-APXAHGVTSAPD 183

Qy     199 AATTPAATTPAA----------TTPAATTPAA----------TTPAATTPGVPAPTSAPV 238
           | :| | |            | :| | |            | :| |     ||||
Db     184 XRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPD 243

Qy     239 WPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSP 298
           |       | |                 ::| |                     :|
Db     244 XRPX-----PGSTAPX-----------------AHGVT-----------------SAP 262

Qy     299 ANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGN 358
```

```
Db    263 DXRPXPGSTAPX-----------------------------------------AHG- 278

Qy    359 LVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP-----TEPIE 413
Db    278 -------------VTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPD 323

Qy    414 WEPLPNG-CPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCPGNLHFSPATQSCESP 469
Db    324 XRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS----------TAP 373

Qy    470 VTAGCQVFECDSDNQCTSTAAPTA-APTAAPTAAP---TAAPTAAPSTVVPPATP-PATA 524
Db    374 XAHG--VTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGST 431

Qy    525 AP-VPPTTAIP--TPAP-TAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEI-PIT 579
Db    432 APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPX---PGSTAPXAHGVTSAPDXRPXP 488

Qy    580 VTSAP-----TAAPTAAP---TAAPTAAPTTAVPEI-PTTVTSPPTAAPTTAA----PAP 626
Db    489 GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXP 548

Qy    627 NTT------VTVPPTAAPTTAAPAPNT-TVTAPPTAAPTTAAPAPNT-TVTVPPTAAP-- 677
Db    549 GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXP 608

Qy    677 -TAAPPT--VAHAPNTTAAPVTT-------TSAPATTPEDDDIDP 711
Db    609 GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAP 653
```

```
RESULT   7
US-08-479-537A-5
; Sequence 5, Application US/08479537A
; Patent No. 5861381
;  GENERAL INFORMATION:
;    APPLICANT:  CHAMBON, Pierre
;    APPLICANT:  KIENY, Marie-Paule
;    APPLICANT:  LATHE, Richard
;    APPLICANT:  HAREUVENI, Mara
;    TITLE OF INVENTION:  PHARMACEUTICAL COMPOSITION FOR THE
;    TITLE OF INVENTION:  TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,537A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 90/13101
;      FILING DATE:  23-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR91/00835
;      FILING DATE:  23-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/039,320
;      FILING DATE:  04-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/403,576
;      FILING DATE:  14-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  017753-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1867 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  128..1727
;      OTHER INFORMATION:  /note= "The amino acids spanning
;      OTHER INFORMATION:  128 to 1727 constitute a repeated region wherein the repea
;      OTHER INFORMATION:  20 amino acids, 17 of which are fixed.  The number of such
;      OTHER INFORMATION:  repeats varies from 1 to 40."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  134
;      OTHER INFORMATION:  /note= "Amino acid 134 is X1 = Xaa
;      OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC,
;      OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  144
;      OTHER INFORMATION:  /note= "Amino acid 144 is Y = Xaa
;      OTHER INFORMATION:  which is the codon for Thr or Asn wherein Thr = ACT, ACC,
;      OTHER INFORMATION:  or ACG; and Asn = AAT or AAC."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  147
;      OTHER INFORMATION:  /note= "Amino acid 147 is X2 = Xaa
;      OTHER INFORMATION:  which is the codon for Pro or Ala wherein Pro = CCT, CCC,
;      OTHER INFORMATION:  or CCG; and Ala = GCT, GCC, GCA, or GCG."
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  1..21
;      OTHER INFORMATION:  /note= "Amino acids 1 to 21 are a
;      OTHER INFORMATION:  21 amino acid precursor sequence."
US-08-479-537A-5
```

```
  Query Match             6.3%;  Score 275.5;  DB 2;  Length 1867;
  Best Local Similarity   23.5%;  Pred. No. 1e-10;
  Matches  166;  Conservative   45;  Mismatches  291;  Indels  203;  Gaps   35;

Qy     93 PGPPAETTQAPATTQAPTTTQAPTT-------TTQAPTTTTQAPTTTTQAPTTTQAPTTT 145
Db     66 PGSGSSTTQGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 125

Qy    146 QAPTTTQAPTT--TTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAP-----TTITQ 198
Db    126 PAPGST-APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGST-APXAHGVTSAPD 183

Qy    199 AATTPAATTPAA----------TTPAATTPAA----------TTPAATTPGVPAPTSAPV 238
Db    184 XRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPD 243

Qy    239 WPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSP 298
Db    244 XRPX-----PGSTAPX-----------------AHGVT-----------------SAP 262

Qy    299 ANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGN 358
Db    263 DXRPXPGSTAPX-----------------------------------------AHG- 278

Qy    359 LVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP-----TEPIE 413
Db    278 -------------VTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPD 323
```

```
Qy    414 WEPLPNG-CPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCPGNLHFSPATQSCESP 469
Db    324 XRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS----------TAP 373

Qy    470 VTAGCQVFECDSDNQCTSTAAPTA-APTAAPTAAP---TAAPTAAPSTVVPPATP-PATA 524
Db    374 XAHG--VTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGST 431

Qy    525 AP-VPPTTAIP--TPAP-TAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEI-PIT 579
Db    432 APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPX---PGSTAPXAHGVTSAPDXRPXP 488

Qy    580 VTSAP-----TAAPTAAP---TAAPTAAPTTAVPEI-PTTVTSPPTAAPTTAA----PAP 626
Db    489 GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXP 548

Qy    627 NTT------VTVPPTAAPTTAAPAPNT-TVTAPPTAAPTTAAPAPNT-TVTVPPTAAP-- 677
Db    549 GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXP 608

Qy    677 -TAAPPT--VAHAPNTTAAPVTT-------TSAPATTPEDDDIDP 711
Db    609 GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAP 653


RESULT   8
US-08-325-267A-2
; Sequence 2, Application US/08325267A
; Patent No. 5585271
;  GENERAL INFORMATION:
;    APPLICANT:  WATARI, JUNJI
;    APPLICANT:  TAKATA, YOSHIHIRO
;    APPLICANT:  OGAWA, MASAHIRO
;    APPLICANT:  PENTTILA, MERJA
;    APPLICANT:  ONNELA, MAIJA-LEENA
;    APPLICANT:  KERANEN, SIRKKA
;    TITLE OF INVENTION:  YEAST AGGLUTINATION GENES AND YEAST
;    TITLE OF INVENTION:  CONTAINING THEM
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVIS HWY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,267A
;      FILING DATE:  18-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP PCT/JP94/00290
;      FILING DATE:  24-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 38871/1993
;      FILING DATE:  26-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2589-023-0XPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 1537 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-325-267A-2


  Query Match             5.6%;  Score 247;  DB 1;  Length 1537;
  Best Local Similarity   23.4%;  Pred. No. 6.4e-09;
  Matches  169;  Conservative   48;  Mismatches  298;  Indels  206;  Gaps   32;

Qy     88 AGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQA 147
Db    261 SNCTVPDP---SNYAVSTT---TTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRT 314

Qy    148 PTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATT 207
Db    315 PTTASTIITTTTEPWNSTFTSTSTELTTVTGTNGVRTDETIIVIRTPTTATTAITT---TE 371

Qy    208 PAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCN 267
Db    372 PWNSTFTSTSTELTTVTGT--------------------NGLPTDETIIVI-------- 403

Qy    268 LFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDL 327
Db    403 ------------RTPTTA---TTAMTTTQPWNDTFTSTSTEMTTVTG------------ 435

Qy    328 LDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTL 387
Db    435 -TNGLPTDETI--------------------IVIRTPTTATTAMTTTQPWDD------TF 467

Qy    388 PGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQ 447
Db    468 TSTSTEMTTV-----TG------------TNGLPTDETI-------------------- 490

Qy    448 TIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPT-------AAPT 500
Db    490 IVIR--------TPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPT 541

Qy    501 AAPTAAPTAAP--STVVPPATPPATAAPVPPTTAIPTP-------APTAAPTAAPTTAAP 551
Db    542 TATTAMTTTQPWNDTFTSTSTEMTT---VTGTNGLPTDETIIVIRTPTTATTAITTTEPW 598

Qy    552 ESPTTVTVPPTAAPTAAPTTAVP-EIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTT 610
Db    599 NS--TFTSTSTELTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTT 656

Qy    611 VTSPPTAAPTTAAPAPNTTVTV--PPTA--APTTAAPAPNT---TVTAPPTAAPTTAAPA 663
Db    657 VTG------TNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPT 710

Qy    664 PNTTVTV-PPTAAPTAAPPTVAHAP-----NTTAAPVTTTSAPATTPEDDD---IDPPLP 714
Db    711 DETIIVIRTPTTATTAMTTT---QPWNDTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTS 767

Qy    715 NDPINPCVEECNVLPW----AHADCDKYWVCDGNNQ-----VLVV---CSEGLQFNPTTK 762
Db    768 EGLISTTTE-----PWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGL---VTTT 819

Qy    763 T 763
Db    820 T 820


RESULT   9
US-08-450-360-2
; Sequence 2, Application US/08450360
; Patent No. 5656457
;  GENERAL INFORMATION:
;    APPLICANT:  Parkes, Deborah Lynn
;    APPLICANT:  Coates, Stephen Ralph
;    TITLE OF INVENTION:  Herpes Simplex Virus Type 2-Glycoprotein G
;    TITLE OF INVENTION:  Proteins and Polypeptides
;    NUMBER OF SEQUENCES:  12
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Leona L. Lauder
;      STREET:  6 Mariposa Court
;      CITY:  Tiburon
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94920
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/450,360
;      FILING DATE:  25-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/129,021
;      FILING DATE:  29-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lauder, Leona L.
;      REGISTRATION NUMBER:  30,863
;      REFERENCE/DOCKET NUMBER:  D-0012.2A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-435-2034
;      TELEFAX:  415-435-0727
;  GENERAL INFORMATION:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-450-360-2


  Query Match             5.4%;  Score 239;  DB 1;  Length 486;
  Best Local Similarity   31.2%;  Pred. No. 6e-09;
  Matches   90;  Conservative   25;  Mismatches   93;  Indels   80;  Gaps   15;

Qy    408 PTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQS-- 466
          |||  :   |                |  |   | | |             | |    |  | ::
Db    256 PTEDVGVLP-----------PHWAPGALDDGPY----------APFPPRPRFRRALRTDP 294

Qy    466 ------CESPVTAGCQVFECDSDNQCTSTAAPTAAP----------TAAPTAAPTAAPTA 509
                :| | | ::     |   ||: :||:||             ||:  |:| ||
Db    295 EGVDPDVRAPRT-GRRLMALTED---TSSDSPTSAPEKTPLPVSATAMAPSVDPSAEPT- 350

Qy    510 APSTVVPP---ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPT 566
          ||:|  ||   ||  || | |  ||: :|         | ||: ||          :|
Db    350 APATTTPPDEMATQAATVAVTPEETAVASP---------PATASVES----------SPL 390

Qy    567 AAPTTAVPEIPITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAP 626
            |  | |    | ||: :|| |   | |||  |       ||:  : |   ||| | |
Db    391 PAAAAATPGAGHTNTSSASAAKTPPTTPAPTTPP-------PTSTHATP--RPTT--PGP 439

Qy    627 NTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPN-TTVTVPPT 673
            ||   | |  |  |: ||  |  :| ||:|   || |:   |:|  |
Db    440 QTTPPGPATPGPVGASAAP--TADSPLTASPPATAPGPSAANVSVAAT 485


RESULT  10
US-08-123-161A-8
; Sequence 8, Application US/08123161A
; Patent No. 5449616
;  GENERAL INFORMATION:
;    APPLICANT:  Campbell, Kevin P.
;    APPLICANT:  Roberds, Steven L.
;    APPLICANT:  Anderson, Richard D.
;    APPLICANT:  Ibraghimov, Oxana B.
;    APPLICANT:  Yang, Bin
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  ME
;      COUNTRY:  USA
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/123,161A
;      FILING DATE:  16-SEP-93
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/946,234
;      FILING DATE:  14-SEP-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M.
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  UIRF89-11A4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (207) 363-0558
;      TELEFAX:  (207) 363-0528
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  895 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-123-161A-8


  Query Match             5.4%;  Score 236.5;  DB 1;  Length 895;
  Best Local Similarity   36.2%;  Pred. No. 1.7e-08;
  Matches   71;  Conservative    9;  Mismatches   95;  Indels   21;  Gaps    6;

Qy    518 ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIP 577
          |||    |  ||||||  |     ||  ||: |   ||    ||   |       || |
Db    316 ATPTPVTAIGPPTTAIQEPPSRIVPT--PTSPAIAPPTETMAPPVRDP-------VPGKP 366

Qy    578 ITVTSAPTAAPTAAPTAAP------TAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVT 631
           |||      |    ||  |       : | ||   :|  |:| |    ||  |  | ||
Db    367 -TVTIRTRGAIIQTPTLGPIQPTRVSEAGTTVPGQIRPTMTIPGYVEPTAVATPPTTTTK 425

Qy    632 VPPTAAPTTAAPAPN--TTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNT 689
          |  : |  | |: :  || |  ||  | |  | | |  |  |  |   ||:|||      |
Db    426 KPRVSTPKPATPSTDSTTTTTRRPTKKPRTPRPVPRVTTKVSITRLETASPPT---RIRT 482

Qy    690 TAAPVTTTSAPATTPE 705
          | : |     |   ||
Db    483 TTSGVPRGGEPNQRPE 498


RESULT  11
US-08-483-278-8
; Sequence 8, Application US/08483278
; Patent No. 5686073
;  GENERAL INFORMATION:
;    APPLICANT:  Campbell, Kevin P.
;    APPLICANT:  Ibraghimov, Oxana B.
;    APPLICANT:  Ervasti, James M.
;    APPLICANT:  Leveille, Cynthia J.
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
```

```
;      CITY:  York Harbor
;      STATE:  ME
;      COUNTRY:  USA
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,278
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/123,161
;      FILING DATE:  16-SEP-93
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M.
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  UIRF89-11A5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (207) 363-0558
;      TELEFAX:  (207) 363-0528
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  895 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-483-278-8


  Query Match             5.4%;  Score 236.5;  DB 1;  Length 895;
  Best Local Similarity   36.2%;  Pred. No. 1.7e-08;
  Matches   71;  Conservative    9;  Mismatches   95;  Indels   21;  Gaps    6;

Qy     518 ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIP 577
           |||     |  ||||||  |     ||  ||: |   ||    ||   |       ||  |
Db     316 ATPTPVTAIGPPTTAIQEPPSRIVPT--PTSPAIAPPTETMAPPVRDP-------VPGKP 366

Qy     578 ITVTSAPTAAPTAAPTAAP------TAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVT 631
            |||      |    ||  |      : | ||   :|  |:| |    ||  |  | ||
Db     367 -TVTIRTRGAIIQTPTLGPIQPTRVSEAGTTVPGQIRPTMTIPGYVEPTAVATPPTTTTK 425

Qy     632 VPPTAAPTTAAPAPN--TTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNT 689
           |  : |  | |: :  || | ||  | | | | |  |  | |   ||:|||      |
Db     426 KPRVSTPKPATPSTDSTTTTTRRPTKKPRTPRPVPRVTTKVSITRLETASPPT---RIRT 482

Qy     690 TAAPVTTTSAPATTPE 705
           | : |    |   ||
Db     483 TTSGVPRGGEPNQRPE 498


RESULT  12
US-08-046-585-5
; Sequence 5, Application US/08046585
; Patent No. 5453362
;  GENERAL INFORMATION:
;    APPLICANT:  Lamarco, Kelly
;    APPLICANT:  Wilson, Angus
;    APPLICANT:  Herr, Winship
;    TITLE OF INVENTION:  A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
;    TITLE OF INVENTION:  HOST CELL FACTOR
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/046,585
;      FILING DATE:  12-APR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-57503-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-046-585-5


  Query Match             5.3%;  Score 232.5;  DB 1;  Length 2035;
  Best Local Similarity   20.4%;  Pred. No. 8e-08;
  Matches  172;  Conservative   71;  Mismatches  302;  Indels  297;  Gaps   34;

Qy      85 AALAGCTLPGP----PAETTQAPATTQA-----PTT----TQAPTTTTQAPTTTT----- 127
           || |    | |    ||   ::||   |     | |    |  |      ||||
Db     406 AATATSPTPNPVPSVPANPPKSPAPAAAAPAVQPLTQVGITLLPQAAPAPPTTTTIQVLP 465

Qy     127 QAPTTTTQAPTTTQAP------TTTQAPTTTQAPTTTTQAPTTTTQAPTTTQA 180
           | ::   ||  :          |    ||  |  | :  :   :|| | |  |   :
Db     466 TVPGSSISVPTAARTQGVPAVLKVTGPQATTGTPLVTMRPASQAGKAPVTVTSLPAGVRM 525

Qy     181 PTTTQAPTTT---QAPTTITQAATTPAATT-----PAATTPAATTPAATTPAATTPGVPA 232
             ||:   |    :|     ||  ||       |::     : || ||   |    |
Db     526 VVPTQSAQGTVIGSSPQMSGMAALAAAAAATQKIPPSSAPTVLSVPAGTTIVKTMAVTPG 585

Qy     233 PTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYV 292
           |: |                                                      |
Db     586 TTTLPA-------------------------------------------------TV 593

Qy     293 QRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYY 352
           :  ||  |        : | |   :     :|  : :                   |   :
Db     594 KVASSPVMV--------SNPATRMLKTAAAQVGTSVSSAT------------NTSTRPI 632

Qy     353 QCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPI 412
             ||     : | |      ||      |  |: || :: : : :           ||
Db     633 ITVH------KSGTVT----VAQQAQ----VVTTVVGGVTKTITLVK----------SPI 668

Qy     413 EWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTI-ARPCPGNLHFSPATQ--SCESP 469
           :| |              |: |: :  |  : :      |     | ||    : |
Db     669 S---VPGGSAL----------ISNLGKVMSVVQTKPVQTSAVTGQASTGPVTQIIQTKGP 715

Qy     470 VTAGCQV-FECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVP 528
           : ||  :     :| : |:    | |  |           : :|||  |  |   |
Db     716 LPAGTILKLVTSADGKPTTIITTTQASGAGTKPTILGISSVSPSTTKPGTT---TIIKTI 772

Qy     529 PTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPT---TAVPEI--------- 577
           | :|| | |     |::|   :|   | :|     : | ||    ||||:|
Db     773 PMSAIITQAGATGVTSSPGIKSP--ITIITTKVMTSGTGAPAKIITAVPKIATGHGQQGV 830

Qy     577 ----------------------------PITVTSAPTAAPTAA---------------- 592
                                       |:||::   |  |
Db     831 TQVVLKGAPGQPGTILRTVPMGGVRLVTPVTVSAVKPAVTTLVVKGTTGVTTLGTVTGTV 890

Qy     592 ----------PTAAPTAAPTTAVPEI---------PTTVTSPPTAAPTTAAPAPNT-TVT 631
                     |:|  | | | :  |          || :|        |||    | |:|
Db     891 STSLAGAGGHSTSASLATPITTLGTIATLSSQVINPTAITVSAAQTTLTAAGGLTTPTIT 950
```

```
Qy    632 VPPTAAPTTAAPAPNTTVTAPP-------------TAAPTTAAPAPNTTVTVPPTAAPT 677
          : | : ||       | :|||                |:|||    |  |||:  :
Db    951 MQPVSQPTQV-----TLITAPSGVEAQPVHDLPVSILASPTT--EQPTATVTIADSGQGD 1003

Qy    678 AAPPTV----AHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHA 733
            | ||    :: |  |    || :|  |   :      |   |    |:
Db   1004 VQPGTVTLVCSNPPCETHETGTTNTATTTVVAN------LGGHPQPTQVQ---------- 1048

Qy    734 DCDKYWVCD--------------GNNQVLVVCSE--------GLQFNPTTKTCDFACNV 770
                :|||              |  |: |||          |     || | : |
Db   1048 -----FVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCETHETGTTNTATTATSNMAGQH 1102

Qy    771 GC 772
          ||
Db   1103 GC 1104


RESULT  13
US-08-393-703-5
; Sequence 5, Application US/08393703
; Patent No. 5585239
;  GENERAL INFORMATION:
;    APPLICANT:  Lamarco, Kelly
;    APPLICANT:  Wilson, Angus
;    APPLICANT:  Herr, Winship
;    TITLE OF INVENTION:  A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
;    TITLE OF INVENTION:  HOST CELL FACTOR
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,703
;      FILING DATE:  24-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-57503-2/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-393-703-5


  Query Match             5.3%;  Score 232.5;  DB 1;  Length 2035;
  Best Local Similarity   20.4%;  Pred. No. 8e-08;
  Matches  172;  Conservative   71;  Mismatches  302;  Indels  297;  Gaps   34;

Qy     85 AALAGCTLPGP----PAETTQAPATTQA-----PTT----TQAPTTTTQAPTTTT----- 127
          || |    | |    ||   ::||   |     | |    |  |      |||||
Db    406 AATATSPTPNPVPSVPANPPKSPAPAAAAPAVQPLTQVGITLLPQAAPAPPTTTTIQVLP 465

Qy    127 QAPTTTTQAPTTTQAP------TTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180
            | ::   ||  :          |    ||  |  | :  :   :|| | |  |   :
Db    466 TVPGSSISVPTAARTQGVPAVLKVTGPQATTGTPLVTMRPASQAGKAPVTVTSLPAGVRM 525

Qy    181 PTTTQAPTTT---QAPTTITQAATTPAATT-----PAATTPAATTPAATTPAATTPGVPA 232
            ||:   |    :|     ||   ||       |::     : || ||   |    |
Db    526 VVPTQSAQGTVIGSSPQMSGMAALAAAAAATQKIPPSSAPTVLSVPAGTTIVKTMAVTPG 585

Qy    233 PTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYV 292
          |: |                                                       |
Db    586 TTTLPA-------------------------------------------------------TV 593

Qy    293 QRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYY 352
          :   || |        :.| |   :     :|  : :               |   :
Db    594 KVASSPVMV--------SNPATRMLKTAAAQVGTSVSSAT------------NTSTRPI 632

Qy    353 QCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPI 412
            ||      : |  |      ||      |  |: || :: : : :           ||
Db    633 ITVH------KSGTVT----VAQQAQ----VVTTVVGGVTKTITLVK----------SPI 668

Qy    413 EWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTI-ARPCPGNLHFSPATQ--SCESP 469
          :| |              |: |: :  |  : :      |     | ||    : |
Db    669 S---VPGGSAL----------ISNLGKVMSVVQTKPVQTSAVTGQASTGPVTQIIQTKGP 715

Qy    470 VTAGCQV-FECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVP 528
          : ||  :     :| : |:    | |  |          : :|||  |  |   |
Db    716 LPAGTILKLVTSADGKPTTIITTTQASGAGTKPTILGISSVSPSTTKPGTT---TIIKTI 772

Qy    529 PTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPT---TAVPEI--------- 577
          | :|| | |     |::|   :|   | :|     : | ||    ||||:|
Db    773 PMSAIITQAGATGVTSSPGIKSP--ITIITTKVMTSGTGAPAKIITAVPKIATGHGQQGV 830

Qy    577 ----------------------------PITVTSAPTAAPTAA---------------- 592
                                       |:||::   |  |
Db    831 TQVVLKGAPGQPGTILRTVPMGGVRLVTPVTVSAVKPAVTTLVVKGTTGVTTLGTVTGTV 890

Qy    592 ----------PTAAPTAAPTTAVPEI---------PTTVTSPPTAAPTTAAPAPNT-TVT 631
                     |:|  | | | :  |          || :|        |||    | |:|
Db    891 STSLAGAGGHSTSASLATPITTLGTIATLSSQVINPTAITVSAAQTTLTAAGGLTTPTIT 950

Qy    632 VPPTAAPTTAAPAPNTTVTAPP-------------TAAPTTAAPAPNTTVTVPPTAAPT 677
          : | : ||       | :|||                |:|||    |  |||:  :
Db    951 MQPVSQPTQV-----TLITAPSGVEAQPVHDLPVSILASPTT--EQPTATVTIADSGQGD 1003

Qy    678 AAPPTV----AHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHA 733
            | ||    :: |  |    || :|  |   :      |   |    |:
Db   1004 VQPGTVTLVCSNPPCETHETGTTNTATTTVVAN------LGGHPQPTQVQ---------- 1048

Qy    734 DCDKYWVCD--------------GNNQVLVVCSE--------GLQFNPTTKTCDFACNV 770
                :|||              |  |: |||          |     || | : |
Db   1048 -----FVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCETHETGTTNTATTATSNMAGQH 1102

Qy    771 GC 772
          ||
Db   1103 GC 1104


RESULT  14
PCT-US93-11721-5
; Sequence 5, Application PC/TUS9311721
;  GENERAL INFORMATION:
;    APPLICANT:  Lamarco, Kelly
;    APPLICANT:  Wilson, Angus
;    APPLICANT:  Herr, Winship
;    TITLE OF INVENTION:  A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
;    TITLE OF INVENTION:  HOST CELL FACTOR
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/11721
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  FP-57503-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
PCT-US93-11721-5


  Query Match             5.3%;  Score 232.5;  DB 3;  Length 2035;
  Best Local Similarity   20.4%;  Pred. No. 8e-08;
  Matches  172;  Conservative   71;  Mismatches  302;  Indels  297;  Gaps   34;

Qy      85 AALAGCTLPGP----PAETTQAPATTQA-----PTT----TQAPTTTTQAPTTTT----- 127
           || |    | |    ||   ::||   |     | |    |  |      |||||
Db     406 AATATSPTPNPVPSVPANPPKSPAPAAAAAPAVQPLTQVGITLLPQAAPAPPTTTTIQVLP 465

Qy     127 QAPTTTTQAPTTTQAP------TTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180
             | ::   ||  :        |    ||  |  | :  :   :|| | |  |   :
Db     466 TVPGSSISVPTAARTQGVPAVLKVTGPQATTGTPLVTMRPASQAGKAPVTVTSLPAGVRM 525

Qy     181 PTTTQAPTTT---QAPTTITQAATTPAATT-----PAATTPAATTPAATTPAATTPGVPA 232
             ||:   |    :|     ||   ||        |::     : || ||   |    |
Db     526 VVPTQSAQGTVIGSSPQMSGMAALAAAAAATQKIPPSSAPTVLSVPAGTTIVKTMAVTPG 585

Qy     233 PTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYV 292
           |: |                                                       |
Db     586 TTTLPA-------------------------------------------------------TV 593

Qy     293 QRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYY 352
           :   ||  |       : | |   :     :|  : :                     |   :
Db     594 KVASSPVMV--------SNPATRMLKTAAAQVGTSVSSAT------------NTSTRPI 632

Qy     353 QCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPI 412
             ||      : |  |       ||       |  |: || :: : : :           ||
Db     633 ITVH------KSGTVT----VAQQAQ----VVTTVVGGVTKTITLVK----------SPI 668

Qy     413 EWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTI-ARPCPGNLHFSPATQ--SCESP 469
           :| |               |: |: :  |  : :      |     | ||    : |
Db     669 S---VPGGSAL----------ISNLGKVMSVVQTKPVQTSAVTGQASTGPVTQIIQTKGP 715

Qy     470 VTAGCQV-FECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVP 528
           : ||  :     :| : |:    | |  |          : :|||  |  |   |
Db     716 LPAGTILKLVTSADGKPTTIITTTQASGAGTKPTILGISSVSPSTTKPGTT---TIIKTI 772

Qy     529 PTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPT---TAVPEI--------- 577
           | :|| | |     |::|   :|   | :|     : | ||    ||||:|
Db     773 PMSAIITQAGATGVTSSPGIKSP--ITIITTKVMTSGTGAPAKIITAVPKIATGHGQQGV 830

Qy     577 ----------------------------PITVTSAPTAAPTAA---------------- 592
                                        |:||::   |  |
Db     831 TQVVLKGAPGQPGTILRTVPMGGVRLVTPVTVSAVKPAVTTLVVKGTTGVTTLGTVTGTV 890

Qy     592 ----------PTAAPTAAPTTAVPEI---------PTTVTSPPTAAPTTAAPAPNT-TVT 631
                    |:|  | | | :  |           || :|       |||    | |:|
Db     891 STSLAGAGGHSTSASLATPITTLGTIATLSSQVINPTAITVSAAQTTLTAAGGLTTPTIT 950

Qy     632 VPPTAAPTTAAPAPNTTVTAPP----------TAAPTTAAPAPNTTVTVPPTAAPT 677
           : | : ||       | :|||              |:|||    |  |||:  :
Db     951 MQPVSQPTQV-----TLITAPSGVEAQPVHDLPVSILASPTT--EQPTATVTIADSGQGD 1003

Qy     678 AAPPTV----AHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHA 733
             | ||    :: |  |    || :|  |   :      |   |    |:
Db    1004 VQPGTVTLVCSNPPCETHETGTTNTATTTVVAN------LGGHPQPTQVQ---------- 1048

Qy     734 DCDKYWVCD--------------GNNQVLVVCSE--------GLQFNPTTKTCDFACNV 770
                :|||                |  |: |||         |     || | : |
Db    1048 -----FVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCETHETGTTNTATTATSNMAGQH 1102

Qy     771 GC 772
           ||
Db    1103 GC 1104


RESULT  15
US-08-078-683A-8
; Sequence 8, Application US/08078683A
; Patent No. 5486599
;  GENERAL INFORMATION:
;    APPLICANT:  Saunders, Scott
;    APPLICANT:  Bernfield, Merton
;    APPLICANT:  Kato, Masato
;    TITLE OF INVENTION:  Construction and USe of Synthetic
;    TITLE OF INVENTION:  Constructs Encoding Syndecan
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII (text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/078,683A
;      FILING DATE:  17-JUN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  CME-062
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  357 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-078-683A-8


  Query Match             5.1%;  Score 224;  DB 1;  Length 357;
  Best Local Similarity   29.8%;  Pred. No. 4.2e-08;
  Matches   81;  Conservative   24;  Mismatches  109;  Indels   58;  Gaps   10;

Qy     472 AGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPAT------------ 520
           :|   || :|      ||       |: |   ||        : | ||
Db      69 SGSGYFEQESG---LETAVSLTTDTSVP--LPTTVAVLPVTLVQPMATPFELFPTEDTSP 123

Qy     520 ---------PPATAAPVPP----TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPT 566
                    |  | ||| |    |||  | : : : |:  || |  : || |   |   |
Db     124 EQTTSVLYIPKITEAPVIPSWKTTTASTTASDSPSTTSTTTTTAATTTTTTTTISTTVAT 183
```

```
Qy     567 AAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPE--IPTTVTS----PPTAAPT 620
           : |||     :|  || | |    |   |       |||::||  : | ||:    | : |
Db     184 SKPTTTQRFLPPFVTKAATTRATTLET------PTTSIPETSVLTEVTTSRLVPSSTAKP 237

Qy     621 TAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAP 680
            : | |:|: |  ||   |     |::  | ||| ||    |   |||       ||::
Db     238 RSLPKPSTSRTAEPTEKSTA---LPSSPTTLPPTEAP-QVEPGELTTVLDSDLEVPTSSG 293

Qy     681 P------------TVAHAPNTTAAPVTTTSAP 700
           |            |     |   | ||  :||
Db     294 PSGDFEIQEEEETTRPELGNEVVAVVTPPAAP 325


Search completed: July 18, 1999, 07:57:25
Job time: 446 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 03:46:46 ; Search time 42.46 Seconds
(without alignments)
923.231 Million cell updates/sec

Title: US-09-294-663-4
Perfect score: 4395
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 807

Scoring table: BLOSUM62

Searched: 396714 seqs, 48575446 residues

Database : Pending_Patents_AA:*
1: /cgn2_6/ptodata/1/paa/PCTUS9_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US084A_COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep:*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/1/paa/US60_COMB.pep:*
19: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep:*
20: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*
21: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*
22: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*
23: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4387	99.8	807	16	US-09-294-663-4	Sequence 4, Appli
2	4352	99.0	805	15	US-09-103-429A-4	Sequence 4, Appli
3	4266.5	97.1	788	16	US-09-294-663-3	Sequence 3, Appli
4	4231.5	96.3	786	15	US-09-103-429A-3	Sequence 3, Appli
5	542	12.3	1404	2	US-07-643-502C-13	Sequence 13, Appl
6	542	12.3	1205	2	US-07-643-502C-28	Sequence 28, Appl
7	542	12.3	1404	2	US-07-757-022B-2	Sequence 2, Appli
8	542	12.3	941	2	US-07-757-022B-14	Sequence 14, Appl
9	542	12.3	1361	2	US-07-757-022B-40	Sequence 40, Appl
10	542	12.3	1311	2	US-07-757-022B-42	Sequence 42, Appl
11	542	12.3	1270	2	US-07-757-022B-44	Sequence 44, Appl
12	542	12.3	1320	2	US-07-757-022B-46	Sequence 46, Appl
13	542	12.3	1354	2	US-07-757-022B-48	Sequence 48, Appl
14	542	12.3	1314	2	US-07-757-022B-50	Sequence 50, Appl
15	542	12.3	1363	2	US-07-757-022B-52	Sequence 52, Appl
16	542	12.3	1049	2	US-07-757-022B-58	Sequence 58, Appl
17	542	12.3	1320	2	US-07-757-022B-60	Sequence 60, Appl
18	542	12.3	1404	2	US-07-757-022B-62	Sequence 62, Appl
19	542	12.3	1038	2	US-07-757-022B-74	Sequence 74, Appl
20	542	12.3	1022	2	US-07-757-022B-84	Sequence 84, Appl
21	542	12.3	1140	2	US-07-757-022B-104	Sequence 104, App
22	542	12.3	1313	2	US-07-757-022B-142	Sequence 142, App
23	463	10.5	1837	13	US-08-928-361-5	Sequence 5, Appli
24	463	10.5	1837	21	US-08-928-361B-5	Sequence 5, Appli
25	453.5	10.3	1721	1	PCT-US97-14104-5	Sequence 5, Appli
26	453.5	10.3	1721	11	US-08-700-651A-5	Sequence 5, Appli
27	453.5	10.3	1721	11	US-08-700-651-5	Sequence 5, Appli
28	453.5	10.3	1721	13	US-08-928-361-6	Sequence 6, Appli
29	453.5	10.3	1721	21	US-08-928-361B-6	Sequence 6, Appli
30	410	9.3	1367	18	US-60-087-236-55	Sequence 55, Appl
31	372	8.5	557	16	US-09-248-796-26892	Sequence 26892, A
32	372	8.5	557	18	US-60-096-409-26892	Sequence 26892, A
33	339	7.7	827	16	US-09-248-796-17307	Sequence 17307, A
34	339	7.7	827	18	US-60-096-409-17307	Sequence 17307, A
35	335	7.6	3880	3	US-08-028-021-1	Sequence 1, Appli
36	332.5	7.6	297	9	US-08-580-545B-6	Sequence 6, Appli
37	332.5	7.6	297	16	US-09-262-653A-6	Sequence 6, Appli
38	325	7.4	907	11	US-08-783-774-2	Sequence 2, Appli
39	309.5	7.0	267	1	PCT-US99-03265-137	Sequence 137, App
40	309.5	7.0	267	1	PCT-US99-03268-142	Sequence 142, App
41	309.5	7.0	267	12	US-08-818-111-137	Sequence 137, App
42	309.5	7.0	267	12	US-08-818-112-142	Sequence 142, App
43	309.5	7.0	216	13	US-08-928-361-8	Sequence 8, Appli
44	309.5	7.0	267	13	US-08-942-341-137	Sequence 137, App
45	309.5	7.0	267	13	US-08-942-578-142	Sequence 142, App

ALIGNMENTS

```
RESULT 1
US-09-294-663-4
; Sequence 4, Application US/09294663
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga Street
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/294,663
;      FILING DATE:  19-APR-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/103,429
;      FILING DATE:  24-JUN-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39-CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  807 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
```

```
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  peritrophic membrane
US-09-294-663-4


  Query Match             99.8%;  Score 4387;  DB 16;  Length 807;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-273;
  Matches  807;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYG 60

Qy     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180

Qy    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240

Qy    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300

Qy    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360

Qy    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420

Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 660

Qy    661 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 720

Qy    721 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 780

Qy    781 ESYEGVQVFIPWNKLDEDIRQALNFEL 807
          |||||||||||||||||||||||||||
Db    781 ESYEGVQVFIPWNKLDEDIRQALNFEL 807


RESULT   2
US-09-103-429A-4
; Sequence 4, Application US/09103429A
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,429A
;      FILING DATE:  24-JUN-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  805 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  peritrophic membrane
US-09-103-429A-4


  Query Match             99.0%;  Score 4352;  DB 15;  Length 805;
  Best Local Similarity   99.8%;  Pred. No. 5.6e-271;
  Matches  805;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYG 60

Qy     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180
          ||||||||  ||||||||||||||||  ||||||||||||||||||||||||||||||||
Db    121 APTTTTQA-TTTTQAPTTTQAPTTTQA-TTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 178

Qy    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 238

Qy    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 298

Qy    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 358

Qy    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 418
```

```
Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 478

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    479 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 538

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 598

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    599 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 658

Qy    661 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    659 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 718

Qy    721 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    719 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 778

Qy    781 ESYEGVQVFIPWNKLDEDIRQALNFEL 807
          |||||||||||||||||||||||||||
Db    779 ESYEGVQVFIPWNKLDEDIRQALNFEL 805


RESULT   3
US-09-294-663-3
; Sequence 3, Application US/09294663
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga Street
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/294,663
;      FILING DATE:  19-APR-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/103,429
;      FILING DATE:  24-JUN-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39-CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  788 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  peritrophic membrane
US-09-294-663-3


  Query Match             97.1%;  Score 4266.5;  DB 16;  Length 788;
  Best Local Similarity   97.1%;  Pred. No. 1.6e-265;
  Matches  784;  Conservative    0;  Mismatches    4;  Indels   19;  Gaps    1;

Qy      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYG 60
          |||||||||||||||||||||||||||||||| ||||| ||| ||||||||||||||
Db      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60

Qy     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180

Qy    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240

Qy    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300

Qy    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360

Qy    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420

Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTV----------------PPTAAPTTA 641

Qy    661 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    642 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 701

Qy    721 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    702 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 761

Qy    781 ESYEGVQVFIPWNKLDEDIRQALNFEL 807
          |||||||||||||||||||||||||||
Db    762 ESYEGVQVFIPWNKLDEDIRQALNFEL 788


RESULT   4
US-09-103-429A-3
; Sequence 3, Application US/09103429A
;  GENERAL INFORMATION:
;    APPLICANT:  Granados, Robert R
```

```
;    APPLICANT:  Wang, Ping
;    TITLE OF INVENTION:  A Novel Invertebrate Intestinal Mucin
;    TITLE OF INVENTION:  cDNA and Related Products and Methods
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brown, Pinnisi & Michaels, P.C.
;      STREET:  118 North Tioga
;      CITY:  Ithaca
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  14850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,429A
;      FILING DATE:  24-JUN-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michaels, Christopher A
;      REGISTRATION NUMBER:  34,390
;      REFERENCE/DOCKET NUMBER:  BTI-39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (607) 256-2000
;      TELEFAX:  (607) 256-3628
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  786 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Trichoplusia ni
;      TISSUE TYPE:  peritrophic membrane
US-09-103-429A-3


  Query Match             96.3%;  Score 4231.5;  DB 15;  Length 786;
  Best Local Similarity   96.9%;  Pred. No. 2.8e-263;
  Matches  782;  Conservative    0;  Mismatches    4;  Indels   21;  Gaps    3;

Qy      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYG 60
          ||||||||||||||||||||||||||||||||| |||||  ||| ||||||||||||
Db      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60

Qy     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQA 180
          |||||||| |||||||||||||||| |||||||||||||||||||||||||||||||||
Db    121 APTTTTQA-TTTTQAPTTTQAPTTTQA-TTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 178

Qy    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 238

Qy    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 298

Qy    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 358

Qy    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 418

Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 478

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    479 SDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 538

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 598

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 660
          ||||||||||||||||||||||||||||||||                ||||||||||
Db    599 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTV----------------PPTAAPTTA 639

Qy    661 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    640 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 699

Qy    721 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    700 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 759

Qy    781 ESYEGVQVFIPWNKLDEDIRQALNFEL 807
          ||||||||||||||||||||||||||
Db    760 ESYEGVQVFIPWNKLDEDIRQALNFEL 786


RESULT   5
US-07-643-502C-13
; Sequence 13, Application US/07643502C
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  A Megakaryocytopoietic Factor
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/643,502C
;      FILING DATE:  19910118
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5166D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  13:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1404 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-643-502C-13


  Query Match             12.3%;  Score 542;  DB 2;  Length 1404;
  Best Local Similarity   30.4%;  Pred. No. 5.3e-27;
  Matches  215;  Conservative   23;  Mismatches  194;  Indels  276;  Gaps   37;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142
           | | || ||: || |      ||| : | :||     |||   ||||  : |||||: ||
Db     331 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 390

Qy     143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192
           || :  |||||: |  ||    ||||| : |||||  |||| :  |||| :  |||| :
Db     391 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 450

Qy     193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
           |     | |||   ||  |||    |    || |||  |||| ||  |
Db     451 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 492

Qy     253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
                                                 |   :  :|   :     |||
Db     492 ------------------------------------PTTPKEPAPTTPK-----EPAPT 509

Qy     313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372
           |:                 |:
Db     510 TTKE---------------PS---------------------------------- 516

Qy     373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429
                                          || |   || |      |
Db     516 ----------------------------------PTTP--KEPAPTTTKSAPT------ 533

Qy     430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489
                         |  | |     :| |    || |              |
Db     533 -----------------TTKEPAPTTTKSAPTTPKEPSPTT-------------TKEP 560

Qy     490 APTA----APTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
           |||     |||     |||     ||:|   || |  |  | |  ||   ||||     |
Db     561 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 617

Qy     546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593
           |||    :|||      | |   |||     |||| || |   |    |||    |||
Db     618 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 675

Qy     594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT----------- 629
                |||||     |||      |:|   ||||           ||| |
Db     676 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 735

Qy     629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683
           | |  ||:  |:  |||     |  | |   ||||  ||| |     ||     || |
Db     736 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 794

Qy     684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720
               || ||  | :|||   || |  |     |  | |  |  |
Db     795 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 842


RESULT   6
US-07-643-502C-28
; Sequence 28, Application US/07643502C
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  A Megakaryocytopoietic Factor
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/643,502C
;      FILING DATE:  19910118
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5166D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1205 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-07-643-502C-28


  Query Match             12.3%;  Score 542;  DB 2;  Length 1205;
  Best Local Similarity   30.7%;  Pred. No. 4.6e-27;
  Matches  212;  Conservative   27;  Mismatches  210;  Indels  242;  Gaps   35;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
           | | || ||: || |      ||| : | :||           ||||  : |||||  ||
Db     132 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 191

Qy     137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
           || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db     192 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 251

Qy     186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
            |||| : ||  |     |    || |||    | |   || |||  ||||: |  |
Db     252 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 309

Qy     244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                     |   :  ||   :
Db     309 ----------------------------------------------PTTTKEPSPTTPK- 322

Qy     304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
               |||  |
Db     322 ----EPAPTTT---------------------------------------------- 329

Qy     364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                       :|      | |  |    || | |  |     |
Db     329 --------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 361

Qy     423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
           |  :     | |                              :| |    :| |
Db     362 APTT-----PKEP-----------------------APTTPKKPAPTT---------- 382

Qy     483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                 |||     |||     ||||   |:   |  | | |   : |||     ||
```

```
Db    382 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 438

Qy    536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589

Db    439 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 496

Qy    590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTT-AAPAPN---TTVTVPPTAAPTTAAPAP 645

Db    497 PAPTTPKETAPTTPKETAPTTLKEP---APTTPKKPAPKELAPTTTKEPTST-TSDKPAP 552

Qy    646 ----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH------APNTTAAPV 694

Db    553 TTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPT 612

Qy    695 TTTS---APATTPEDDDIDP--PLPNDPINP 720

Db    613 STTSDKPAPTTPKETAPTTPKEPAPTTPKKP 643


RESULT   7
US-07-757-022B-2
; Sequence 2, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1404 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-2


  Query Match            12.3%;  Score 542;  DB 2;  Length 1404;
  Best Local Similarity  30.4%;  Pred. No. 5.3e-27;
  Matches  215;  Conservative   23;  Mismatches  194;  Indels  276;  Gaps   37;

Qy     93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142

Db    331 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 390

Qy    143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192

Db    391 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 450

Qy    193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252

Db    451 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 492

Qy    253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312

Db    492 ----------------------------------PTTPKEPAPTTPK-----EPAPT 509

Qy    313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372

Db    510 TTKE---------------PS---------------------------------- 516

Qy    373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429

Db    516 ---------------------------------PTTP--KEPAPTTTKSAPT------ 533

Qy    430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489

Db    533 -----------------TTKEPAPTTTKSAPTTPKEPSPTT--------------TKEP 560

Qy    490 APTA----APTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545

Db    561 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 617

Qy    546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593

Db    618 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 675

Qy    594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT---------- 629

Db    676 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 735

Qy    629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683

Db    736 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 794

Qy    684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720

Db    795 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 842


RESULT   8
US-07-757-022B-14
; Sequence 14, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  941 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-14


  Query Match           12.3%;  Score 542;  DB 2;  Length 941;
  Best Local Similarity  30.4%;  Pred. No. 3.6e-27;
  Matches  215;  Conservative   23;  Mismatches  194;  Indels  276;  Gaps   37;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142
           | | || ||: || |      ||| : | :||     |||   ||||  : |||||: ||
Db     132 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 191

Qy     143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192
           || :  |||||: |  ||     ||||| : |||||  ||||  :  ||||  :  |||| :
Db     192 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 251

Qy     193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
           |     | |||   ||  |||    |    || |||  |||| ||  |
Db     252 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 293

Qy     253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
                                                |  :  :|  :      |||
Db     293 ------------------------------------PTTPKEPAPTTPK-----EPAPT 310

Qy     313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372
           |:                  |:
Db     311 TTKE---------------PS---------------------------------------- 317

Qy     373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429
                                          || |   || |      |
Db     317 ----------------------------------PTTP--KEPAPTTTKSAPT------ 334

Qy     430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489
                         |   | |     :| |    || |             |
Db     334 -----------------TTKEPAPTTTKSAPTTPKEPSPTT-------------TKEP 361

Qy     490 APTA----APTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
           |||     |||     |||      ||:|   || |  |  | |  ||   ||||     |
Db     362 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 418

Qy     546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593
           |||    :|||      | |   |||     ||||  || |   |    |||   |||
Db     419 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 476

Qy     594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT---------- 629
               ||||     |||      |:|   ||||           ||| |
Db     477 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 536

Qy     629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683
           | |  ||:  |:  |||      |  | |   ||||  ||| |     ||     || |
Db     537 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 595

Qy     684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720
               || ||  | :|||   || |  |     |  | |  |  |
Db     596 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 643


RESULT  9
US-07-757-022B-40
; Sequence 40, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1361 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-40


  Query Match           12.3%;  Score 542;  DB 2;  Length 1361;
  Best Local Similarity  30.4%;  Pred. No. 5.2e-27;
  Matches  215;  Conservative   23;  Mismatches  194;  Indels  276;  Gaps   37;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142
```

```
          | | || ||: || |      ||| : | :||     |||   ||||  : |||||: ||
Db    288 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 347

Qy    143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192
          || :  |||||: |  ||    |||||  : |||||  |||| :  |||| :  |||| :
Db    348 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 407

Qy    193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
          |     | |||   ||  |||     |     || |||  |||| ||  |
Db    408 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 449

Qy    253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
                                                |   :  :|  :      |||
Db    449 ------------------------------------PTTPKEPAPTTPK-----EPAPT 466

Qy    313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372
           |:              |:
Db    467 TTKE----------------PS---------------------------------- 473

Qy    373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429
                                          || |   || |      |
Db    473 ----------------------------------PTTP--KEPAPTTTKSAPT------ 490

Qy    430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489
                          |   | |     :| |    || |               |
Db    490 -----------------TTKEPAPTTTKSAPTTPKEPSPTT-------------TKEP 517

Qy    490 APTA----APTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
          |||     |||     |||     ||:|   || |  |  | |  ||   ||||     |
Db    518 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 574

Qy    546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593
          |||    :|||      | |   |||     ||||  || |   |    |||    |||
Db    575 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 632

Qy    594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT---------- 629
                ||||     |||      |:|   ||||           ||| |
Db    633 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 692

Qy    629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683
          | |  ||:  |:  |||     |  | |   ||||  ||| |     ||     || |
Db    693 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 751

Qy    684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720
                  || ||  | :|||   || |  |     |  | |  |  |
Db    752 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 799


RESULT 10
US-07-757-022B-42
; Sequence 42, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1311 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-42


  Query Match             12.3%;  Score 542;  DB 2;  Length 1311;
  Best Local Similarity   30.4%;  Pred. No. 5e-27;
  Matches  215;  Conservative   23;  Mismatches  194;  Indels  276;  Gaps   37;

Qy     93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142
          | | || ||: || |      ||| : | :||     |||   ||||  : |||||: ||
Db    238 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 297

Qy    143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192
          || :  |||||: |  ||    |||||  : |||||  |||| :  |||| :  |||| :
Db    298 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 357

Qy    193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
          |     | |||   ||  |||     |     || |||  |||| ||  |
Db    358 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 399

Qy    253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
                                                |   :  :|  :      |||
Db    399 ------------------------------------PTTPKEPAPTTPK-----EPAPT 416

Qy    313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372
           |:              |:
Db    417 TTKE----------------PS---------------------------------- 423

Qy    373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429
                                          || |   || |      |
Db    423 ----------------------------------PTTP--KEPAPTTTKSAPT------ 440

Qy    430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489
                          |   | |     :| |    || |               |
Db    440 -----------------TTKEPAPTTTKSAPTTPKEPSPTT-------------TKEP 467

Qy    490 APTA----APTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
          |||     |||     |||     ||:|   || |  |  | |  ||   ||||     |
Db    468 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 524

Qy    546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593
          |||    :|||      | |   |||     ||||  || |   |    |||    |||
Db    525 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 582

Qy    594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT---------- 629
                ||||     |||      |:|   ||||           ||| |
Db    583 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 642
```

```
Qy     629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683
           | |  ||:  |:  |||     |  | |  ||||  ||| |     ||     || |
Db     643 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 701

Qy     684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720
                  || ||  | :|||   || |  |     |  | |  |  |
Db     702 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 749


RESULT  11
US-07-757-022B-44
; Sequence 44, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1270 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-44


  Query Match             12.3%;  Score 542;  DB 2;  Length 1270;
  Best Local Similarity   30.4%;  Pred. No. 4.8e-27;
  Matches  215;  Conservative   23;  Mismatches  194;  Indels  276;  Gaps   37;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142
           | | || ||: || |      ||| : | :||     |||   ||||  : |||||: ||
Db     197 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 256

Qy     143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192
           || :  |||||: |  ||    ||||| : |||||  ||||  :  ||||  :  ||||  :

Db     257 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 316

Qy     193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
           |      | |||   ||  |||     |    || |||  |||| ||  |
Db     317 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 358

Qy     253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
                                                 |   :  :|   :      |||
Db     358 ------------------------------------PTTPKEPAPTTPK-----EPAPT 375

Qy     313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372
           |:                 |:
Db     376 TTKE---------------PS------------------------------------ 382

Qy     373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429
                                                 || |   || |      |
Db     382 ----------------------------------PTTP--KEPAPTTTKSAPT------ 399

Qy     430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489
                         |   | |     :| |    || |              |
Db     399 ----------------TTKEPAPTTTKSAPTTPKEPSPTT-------------TKEP 426

Qy     490 APTA----APTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
           |||     |||     |||     ||:|   || | |  | | |  ||  ||||     |
Db     427 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 483

Qy     546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593
           |||    :|||      | |   |||     ||||  || |   |    |||    |||
Db     484 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 541

Qy     594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT---------- 629
               ||||     |||      |:|   ||||           ||| |
Db     542 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 601

Qy     629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683
           | |  ||:  |:  |||     |  | |  ||||  ||| |     ||     || |
Db     602 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 660

Qy     684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720
                  || ||  | :|||   || |  |     |  | |  |  |
Db     661 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 708


RESULT  12
US-07-757-022B-46
; Sequence 46, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1320 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-46


  Query Match             12.3%;  Score 542;  DB 2;  Length 1320;
  Best Local Similarity   30.4%;  Pred. No. 5e-27;
  Matches  215;  Conservative   23;  Mismatches  194;  Indels  276;  Gaps   37;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142
           | | || ||: ||. |      ||| : | :||     |||   ||||  : |||||: ||
Db     247 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 306

Qy     143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192
           || :  |||||: |  ||    ||||| : ||||| |||| :  |||| :  |||| :
Db     307 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 366

Qy     193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
           |     | |||   ||  |||    |    || |||  |||| ||  |
Db     367 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 408

Qy     253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
                                              |   :  :|   :      |||
Db     408 ------------------------------------PTTPKEPAPTTPK-----EPAPT 425

Qy     313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372
           |:                  |:
Db     426 TTKE---------------PS---------------------------------- 432

Qy     373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429
                                             || |   || |      |
Db     432 ----------------------------------PTTP--KEPAPTTTKSAPT------ 449

Qy     430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489
                       |   | |     :| |    || |              |
Db     449 -----------------TTKEPAPTTTKSAPTTPKEPSPTT-------------TKEP 476

Qy     490 APTA----APTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
           |||     |||     |||     ||:|   || |  |  | |  ||   ||||     |
Db     477 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 533

Qy     546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593
           |||    :|||     | |   |||     |||| || |   |    |||    |||
Db     534 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 591

Qy     594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT----------- 629
                |||     |||       |:|   ||||           ||| |
Db     592 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 651

Qy     629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683
           | |  ||:  |: |||     |  | |   ||||   ||| |     ||     || |
Db     652 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 710

Qy     684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720
                   || ||  | :|||   || |  |     |  | |  |  |
Db     711 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 758


RESULT  13
US-07-757-022B-48
; Sequence 48, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1354 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-48


  Query Match             12.3%;  Score 542;  DB 2;  Length 1354;
  Best Local Similarity   30.4%;  Pred. No. 5.1e-27;
  Matches  215;  Conservative   23;  Mismatches  194;  Indels  276;  Gaps   37;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142
           | | || ||: || |      ||| : | :||     |||   ||||  : |||||: ||
Db     281 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 340

Qy     143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192
           || :  |||||: |  ||    ||||| : ||||| |||| :  |||| :  |||| :
Db     341 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 400

Qy     193 PTTITQAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
           |     | |||   ||  |||    |    || |||  |||| ||  |
Db     401 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 442
```

```
Qy    253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
                                          |   :  :|   :      |||
Db    442 ------------------------------------PTTPKEPAPTTPK-----EPAPT 459

Qy    313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372
           |:                 |:
Db    460 TTKE---------------PS---------------------------------- 466

Qy    373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429
                                           || |   || |      |
Db    466 ---------------------------------PTTP--KEPAPTTTKSAPT------ 483

Qy    430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489
                         |   | |     :| |    || |              |
Db    483 -----------------TTKEPAPTTTKSAPTTPKEPSPTT-------------TKEP 510

Qy    490 APTA----APTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
          |||     |||     |||     ||:|   || |  |  | |  ||   ||||     |
Db    511 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 567

Qy    546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593
          |||    :|||      | |   |||     ||||  || |   |    |||    |||
Db    568 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 625

Qy    594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT---------- 629
                |||     |||      |:|   ||||           ||| |
Db    626 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 685

Qy    629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683
          | |  ||:  |: |||     |  | |   ||||  ||| |    ||     || |
Db    686 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 744

Qy    684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720
                ||  ||  | :|||   || |  |     |  | |  |  |
Db    745 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 792
```

```
RESULT  14
US-07-757-022B-50
; Sequence 50, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  50:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1314 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-50


  Query Match             12.3%;  Score 542;  DB 2;  Length 1314;
  Best Local Similarity   30.4%;  Pred. No. 5e-27;
  Matches  215;  Conservative   23;  Mismatches  194;  Indels  276;  Gaps   37;

Qy     93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142
          | | || ||: || |      ||| : | :||     |||   ||||  : |||||: ||
Db    241 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 300

Qy    143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192
          || :  |||||: |  ||    ||||| : |||||  |||| :  |||| :  |||| :
Db    301 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 360

Qy    193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
          |     | |||   ||  |||    |    || |||  |||| ||  |
Db    361 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 402

Qy    253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
                                          |   :  :|   :      |||
Db    402 ------------------------------------PTTPKEPAPTTPK-----EPAPT 419

Qy    313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372
           |:                 |:
Db    420 TTKE---------------PS---------------------------------- 426

Qy    373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429
                                           || |   || |      |
Db    426 ---------------------------------PTTP--KEPAPTTTKSAPT------ 443

Qy    430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489
                         |   | |     :| |    || |              |
Db    443 -----------------TTKEPAPTTTKSAPTTPKEPSPTT-------------TKEP 470

Qy    490 APTA----APTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
          |||     |||     |||     ||:|   || |  |  | |  ||   ||||     |
Db    471 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 527

Qy    546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593
          |||    :|||      | |   |||     ||||  || |   |    |||    |||
Db    528 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 585

Qy    594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT---------- 629
                |||     |||      |:|   ||||           ||| |
Db    586 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 645

Qy    629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683
          | |  ||:  |: |||     |  | |   ||||  ||| |    ||     || |
Db    646 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 704

Qy    684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720
                ||  ||  | :|||   || |  |     |  | |  |  |
Db    705 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 752


RESULT  15
```

```
US-07-757-022B-52
; Sequence 52, Application US/07757022B
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1363 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-52


  Query Match           12.3%;  Score 542;  DB 2;  Length 1363;
  Best Local Similarity  30.4%;  Pred. No. 5.2e-27;
  Matches  215;  Conservative   23;  Mismatches  194;  Indels  276;  Gaps   37;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142
           | | || ||: || |      ||| : | :||     |||   ||||  : |||||: ||
Db     290 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 349

Qy     143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192
           || :  |||||: |  ||    ||||| : |||||  |||| :  |||| :  |||| :
Db     350 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 409

Qy     193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
           |     | ||   ||  |||    |    || |||  |||| ||  |
Db     410 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 451

Qy     253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
                                          |   :  :|  :      |||
Db     451 ------------------------------------PTTPKEPAPTTPK-----EPAPT 468

Qy     313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372
           |:                 |:
Db     469 TTKE---------------PS------------------------------------ 475

Qy     373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429
                                              || |   || |      |
Db     475 ----------------------------------PTTP--KEPAPTTTKSAPT------ 492

Qy     430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489
                         |   | |     :| |    || |              |
Db     492 -----------------TTKEPAPTTTKSAPTTPKEPSPTT-------------TKEP 519

Qy     490 APTA----APTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
           |||     |||     |||     ||:|   || |  |  | |  ||   ||||     |
Db     520 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 576

Qy     546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593
           |||    :|||      | |   |||     ||||  || |   |    |||    |||
Db     577 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 634

Qy     594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT----------- 629
                |||      |||      |:|   ||||           ||| |
Db     635 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 694

Qy     629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683
           | |  ||:  |:  |||     |  | |   ||||  ||| |     ||     || |
Db     695 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 753

Qy     684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720
                   || ||  | :|||   || |  |     |  | |  |  |
Db     754 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 801


Search completed: July 18, 1999, 03:46:52
Job time: 1536 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 00:45:05 ; Search time 27.5 Seconds
(without alignments)
1099.291 Million cell updates/sec

Title: US-09-294-663-4
Perfect score: 4395
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 807

Scoring table: BLOSUM62

Searched: 116738 seqs, 37460341 residues

Database : PIR_58:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	660.5	15.0	3020	2	A43932	mucin 2 precursor,
2	460	10.5	873	2	A47283	calphotin - fruit
3	457	10.4	865	2	A47282	calcium-binding pr
4	446.5	10.2	797	1	VGBEX1	glycoprotein X pre
5	430.5	9.8	662	2	A45155	mucin FIM-C.1 - Af
6	428	9.7	1335	2	A35887	mucin precursor, p
7	428	9.7	1255	2	B35175	episialin B - huma
8	428	9.7	1295	2	A35886	polymorphic epithe
9	428	9.7	1264	2	A35175	episialin A - huma
10	421.5	9.6	1188	2	S49915	extensin-like prot
11	410	9.3	1367	2	S48478	glucan 1,4-alpha-g
12	384	8.7	294	2	A37232	mucin, tracheal (A
13	374	8.5	822	2	A38420	antifreeze glycopr
14	373	8.5	1118	2	A48292	mucin, tracheobron
15	367.5	8.4	474	2	S15921	protein TPX-VT3 -
16	365	8.3	353	2	S36438	EPPT protein - hyd
17	352	8.0	670	2	F36791	hypothetical prote
18	348.5	7.9	279	2	S53363	mucin 5AC (clone J
19	347.5	7.9	350	2	S22456	hydroxyproline-ric
20	345.5	7.9	528	2	I47141	gastric mucin (clo
21	345	7.8	232	2	A60095	larval glue protei
22	341	7.8	303	2	S28264	hydroxyproline-ric
23	340.5	7.7	328	2	JQ0985	hydroxyproline-ric
24	340.5	7.7	216	2	I51920	mucin - rhesus mac
25	337	7.7	901	2	A49227	sialidase - Actino
26	336.5	7.7	267	2	S08314	cell wall glycopro
27	331	7.5	400	2	A28172	spasmolysin precur
28	324.5	7.4	379	2	S50125	larval glue protei
29	324	7.4	1630	2	A53577	ascites sialoglyco
30	321	7.3	907	1	QQBE21	membrane antigen g
31	321	7.3	376	2	S71558	probable cell wall
32	320.5	7.3	377	2	A48018	mucin 7 precursor,
33	318.5	7.2	860	2	JC4566	chitinase (EC 3.2.
34	317	7.2	307	1	GSFF3	salivary glue prot
35	315	7.2	1161	2	S57180	probable membrane
36	313.5	7.1	886	2	S29605	glycoprotein 350/2
37	311	7.1	620	2	S06733	hydroxyproline-ric
38	309.5	7.0	548	2	E70546	hypothetical prote
39	309.5	7.0	327	2	S20074	promastigote surfa
40	309	7.0	369	2	S20500	hydroxyproline-ric
41	308.5	7.0	283	2	S13383	hydroxyproline-ric
42	308.5	7.0	451	2	S71754	cellular hepatitis
43	308	7.0	322	2	A53715	apomucin precursor
44	306.5	7.0	447	2	A39321	mucin - rat (fragm
45	300.5	6.8	3149	1	QQBE8	BPLF1 protein - hu

ALIGNMENTS

RESULT 1
A43932
mucin 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin SMUC-41
C;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 20-Mar-1998
C;Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A;Reference number: A49963
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU1>
A;Cross-references: GB:L21998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A;Reference number: A45106; MUID:93016075
A;Accession: A45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 626-1895 <GU2>
A;Cross-references: GB:M94131; NID:g186395; PID:g186396
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)
A;Accession: B45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2037-3020 <GU3>
A;Cross-references: GB:M94132; NID:g186397; PID:g186398
A;Experimental source: colon
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A;Reference number: A43932; MUID:91358717
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350,'L',1352-1411,'S',1413-1448,'P',1450-1503,'T',1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:g188863; PID:g188864
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A;Reference number: A33532; MUID:89197956
A;Accession: B33532
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:g188873; PID:g188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
J. Clin. Invest. 87, 77-82, 1991
A;Title: Human bronchus and intestine express the same mucin gene.
A;Reference number: A61257
A;Accession: A61257
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>
A;Experimental source: bronchus
R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstn
Biochem. Biophys. Res. Commun. 183, 821-828, 1992

```
A;Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t
A;Reference number: PQ0328; MUID:92198477
A;Accession: PQ0328
A;Molecule type: mRNA
A;Residues: 2328-2468 <XUG>
A;Cross-references: GB:M86523
A;Experimental source: small intestine
A;Accession: PQ0329
A;Molecule type: protein
A;Residues: 2328-2342,'K',2344-2354 <XUG1>
C;Genetics:
A;Gene: GDB:MUC2
A;Cross-references: GDB:120203; OMIM:158370
A;Map position: 11p15.5-11p15.5
C;Superfamily: von Willebrand factor type C repeat homology
C;Keywords: glycoprotein; intestine; tandem repeat
F;2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>


  Query Match             15.0%;  Score 660.5;  DB 2;  Length 3020;
  Best Local Similarity   26.1%;  Pred. No. 1.5e-25;
  Matches  210;  Conservative   41;  Mismatches  204;  Indels  349;  Gaps   27;

Qy      29 LHEPHP-------DXPPAEQXXLLPXEYDCTKF-------------YYCEYGLKFIAPRD 68
           ::| ||       |  | :     | : :|                 |:  : ||   :
Db    1309 INEDHPSSGSDDGDREPFDGVCGAPEDIECRSVKDPHLSLEQHGQKVQCDVSVGFICKNE 1368

Qy      69 CAPGT-------EFKFSAQTCVHAALAGC-TLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
             |        ::|    |    :  | | | ||  |   | ||   |||  ||||
Db    1369 DQFGNGPFGLCYDYKIRVNCC--WPMDKCITTPSPPTTTPSPPPTT---TTTLPPTTTPS 1423

Qy     121 APTTTTQAPTTTTQAPTTTQAP----TTTQAPTTTQAP---TTTTQAPTTTTQAPTTTTQ 173
            |||     ||||  |||| :|    |||  |||| :|   ||||  ||||   ||||
Db    1424 PPTT-----TTTTPPPTTTPSPPITTTTTPLPTTTPSPPISTTTTPPPTTTPSPPTTTPS 1478

Qy     174 APTTTQAP----TTTQAPTTTQAPTTITQAATTPAATT--PAATTPAATTPAATTPAATT 227
            |||| :|    |||  |||| :|  :|   | ||:||  |  |||:  |   |||  ||
Db    1479 PPTTTPSPPTTTTTTPPPTTTPSP-PMTTPITPPASTTTLPPTTTPSPPTTTTTTPPPTT 1537

Qy     228 PGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLY 287
              | ||: |: ||
Db    1538 TPSP-PTTTPITPP---------------------------------------------- 1551

Qy     288 FNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNR 347

Db    1551 ------------------------------------------------------------ 1551

Qy     348 CDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYC 407

Db    1551 ------------------------------------------------------------ 1551

Qy     408 PTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCE 467

Db    1551 ------------------------------------------------------------ 1551

Qy     468 SPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPV 527
                          |:|  ||  |:  ||   |  ||  ||       || |  |
Db    1551 ---------------TSTTTLPPTTTPSPPPTTTTTPPPTTTPS-------PPTTTTPS 1587

Qy     528 PPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAA 587
           |||    || ||       || :| : || | |||  |:   ||     |||    ||:
Db    1588 PPTITTTTPPPT-------TTPSPPTTTTTTPPPTTTPSPPTTT-----PIT---PPTST 1632

Qy     588 PTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAP----TTAAP 643
            |  ||  |:  |||     |||  ||||   || :|   || | |||  |    ||
Db    1633 TTLPPTTTPSPPPTTTTTPPPTTTPSPPTT--TTPSPPITTTTTPPPTTTPSSPITTTPS 1690

Qy     644 APNTTVTAP-PTAAP-----TTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNT--TAAPVT 695
            | ||:| | ||  |     ||  |:  || : |||   | :| |    | |  |  | |
Db    1691 PPTTTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPTTMTTPSPTTTPSPPTTTMTTLPPT 1750

Qy     696 TTSAP-ATTPEDDDIDPPL-----PNDPINPCVEECNVLPWAHADCDKYWVCDGNNQVL- 749
           |||:|  |||    | ||         |  |||  ||   |  :    :    |: :::

Db    1751 TTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDTELIG 1810

Qy     749 VVCSEGLQFNPTTKTCDFACNVGC 772
               ||  |           :| |: |
Db    1811 DVCGPG-----------WAANISC 1823


RESULT   2
A47283
calphotin - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 10-Sep-1997
C;Accession: A47283
R;Ballinger, D.G.; Xue, N.; Harshman, K.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 1536-1540, 1993
A;Title: A Drosophila photoreceptor cell-specific protein, calphotin, binds calcium a
A;Reference number: A47283; MUID:93165730
A;Contents: photoreceptor cells
A;Accession: A47283
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-873 <BAL>
A;Cross-references: GB:L05080; NID:g157071; PID:g157072
A;Note: sequence extracted from NCBI backbone (NCBIN:124958, NCBIP:124959)


  Query Match             10.5%;  Score 460;  DB 2;  Length 873;
  Best Local Similarity   24.1%;  Pred. No. 3.8e-16;
  Matches  198;  Conservative   45;  Mismatches  251;  Indels  326;  Gaps   35;

Qy      91 TLPGPPAETTQAPAT---TQAPTTTQAPTTTTQAPTTTTQAP----------------- 130
           |:| | :    || |    ||    :|     ||     ||
Db       5 TIPSPVSAPVAAPVTPSAVAAPVQVVSPAAVAVAPAVVAPAPAAPTAVTPVAPPPTLASV 64

Qy     130 ---TTTTQAP----TTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPT 182
             | |  ||      : ||  : ||    |||    :| :|   |    | |    ||
Db      65 QPATVTVPAPAPIAAASVAPVASVAPPVVAAPTPPAASPVST--PPVAVAQIPVAVSAPV 122

Qy     183 TTQAPTTTQAPTTITQ-AATTPAATTP----AATTPAATTPAATTPAATTPGVPAPTSAP 237
             ||     || :       |   ||    :| |||| ||  :   || | ||| |: |
Db     123 ---APPVAATPTPVAPIPVAAPVIATPPVAASAPTPAAVTPVVSPVIATPPVVPANTTVP 179

Qy     238 VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDS 297
           |  |:   :|   |                                  | :      :
Db     180 VAAPVA-AVPAAVPV-----------------------------------VAPVLAPAVA 203

Qy     298 PANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHG 357
           ||      |    ||| |      | :  : :   |       |:|  :
Db     204 PAVAPVVAETPAPPPVAE------IPVATIPECVAP-------LIPEVS----------- 240

Qy     358 NLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPL 417
                                   |  | |   :| | |   |       ||
Db     240 -------------------------VVATKPLAAAEPVVVAPPA-------TE------T 261

Qy     418 PNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVF 477
           |   ||  |    || |                |      :| : | | || |
Db     262 PVVAPAAAS-----PHVS--------------VAPAVETAVVAPVSASTEPPVAA----- 298

Qy     478 ECDSDNQCTSTAAPTAAPTAAPTAA-------------PTAAPTAAPSTVVPP------- 518
                 | | ||   |  ||  |              || ||   | |: ||
Db     298 -------ATLTTAP-ETPALAPVVAESQVAANTVVATPPTPAP--EPETIAPPVVAETPE 347

Qy     518 --------ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVT-----VPPTAA 564
                    |||     |||  |   |||  | |  | | |   |          :||  |
Db     348 VASVAVAETTPPV----VPPVAAESIPAPVVATTPVPATLAVTDPDVTASAVPELPPVIA 403

Qy     565 PTAAPTTAVPEIPITV----------------------TSAPTAAPT------------ 590
           |:  | :|| | |: :                     | || :||
Db     404 PSPVP-SAVAETPVDLAPPVLPPVAAEPVPAVVAEETPETPAPASAPVTIAALDIPEVAP 462

Qy     590 --AAPTAAPTAAPTTAVPEI---PTTVTSPPTAAPTTAAP----------------AP 626
             |||: ||  ||: | | :   ||| : | | ||  | |                 ||
Db     463 VIAAPSDAPAEAPSAAAPIVSTPPTTASVPETTAPPAAVPTEPIDVSVLSEAAIETPVAP 522
```

```
Qy    627 NTTVT--------VPPTAA-----------------------------------PTTA 641
          ||         || ||                                     | |:
Db    523 PVEVTTEVAVADVAPPEAAEDLIIEPVEPPAPIPDLLEQTTSVPAVEAAESTSSPIPETS 582

Qy    642 APAPNTTVTAPPTA-APTTA---APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPV--T 695
          | ||  | :|  | || ||    | |  ::  |    |  ||  :  | :    ||  |
Db    583 LPPPNEAVASPEVAVAPITAPEPIPEPEPSLATPTEPIPVEAPVVIQEAVDAVEVPVTET 642

Qy    696 TTSAPATTPE------DDDIDPPLPNDPI---NPCVEECN 726
          :|| | || |      :  :|| :   |:    | |   |
Db    643 STSIPETTVEFPEAVAEKVLDPAITEAPVTTQEPDVANIN 682


RESULT   3
A47282
calcium-binding protein calphotin - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C;Accession: A47282
R;Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 1531-1535, 1993
A;Title: Calphotin: a Drosophila photoreceptor cell calcium-binding protein.
A;Reference number: A47282; MUID:93165729
A;Accession: A47282
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-865 <MAR>
A;Cross-references: GB:L02111; NID:g157031; PID:g157032
A;Experimental source: photoreceptor cells
A;Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBIP:124956)
C;Genetics:
A;Gene: FlyBase:Cpn
A;Cross-references: FlyBase:FBgn0010218
C;Keywords: calcium binding


  Query Match             10.4%;  Score 457;  DB 2;  Length 865;
  Best Local Similarity   24.2%;  Pred. No. 5.3e-16;
  Matches  197;  Conservative   46;  Mismatches  250;  Indels  320;  Gaps   35;

Qy     91 TLPGPPAETTQAPAT---TQAPTTTQAPTTTTQAPT-------------TTTQAPTTTT 133
          |:| | :    || |     ||    :|     ||              : |  | |
Db      5 TIPSPVSAPVAAPVTPSAVAAPVQVVSPAAVAPAPAAPIAVTPVAPPPTLASVQPATVTI 64

Qy    134 QAPTTTQA----PTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTT 189
          ||    |    | : ||    |||    :| :|   |    | |    ||    ||
Db     65 PAPAPIAAASVTPVASVAPPVVAAPTPPAASPVST---PVAVAQIPVAVSAPV---APPV 118

Qy    190 TQAPTTITQ-AATTPAATTP----AATTPAATTPAATTPAATTPGVPAPTSAPVWPPICE 244
             || : |     |   ||    :| |||| ||  :   |: | ||| |: ||  |:
Db    119 AATPTPVVQIPVAAPVIATPPVAASAPTPAAVTPVISPVIASPPVVPANTTVPVAAPVA- 178

Qy    245 LLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECD 304
          :|   |                                    | :    :||
Db    178 AVPAAVPV-----------------------------------VAPVLAPAVAPAVAPVV 202

Qy    305 GEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRC 364
          |    ||| |      | :  : :   |       |:|  :
Db    203 AETPAPPPVAE------IPVATIPECVAP-------LIPEVS----------------- 232

Qy    365 GAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPAD 424
                        |  | |   :| | |   |       ||       |   ||
Db    232 ------------------VVATKPLAAAEPVVVAPPA-------TE------TPVVAPAA 260

Qy    425 FSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQ 484
          |     || |                |       :| : | | || |
Db    261 AS-----PHVS--------------VAPAVETAVVAPVSASTEPPVAA------------ 290

Qy    485 CTSTAAPTAAPTAAPTAA-------------PTAAPTAAPSTVVPP-------------- 518
          | | ||   |  ||  |             || ||   | |: ||
Db    290 ATLTTAP-ETPALAPVVAESQVAANTVVATPPTPAP--EPETIAPPVVAETPEVASVAVA 346

Qy    518 -ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVT-----VPPTAAPTAAPTT 571
           |||     |||  |   |||  | |  | | |   |          :||  ||:  | :
Db    347 ETTPPV----VPPVAAESIPAPVVATTPVPATLAVTDPDVTASAVPELPPVIAPSPVP-S 401

Qy    572 AVPEIPITV----------------------TSAPTAAPT--------------AAPTA 594
          || | |: :                      | || :||               |||:
Db    402 AVAETPVDLAPPVLPPVAAEPVPAVVAEETPETPAPASAPVTIAALDIPEVAPVIAAPSD 461

Qy    595 APTAAPTTAVPEI---PTTVTSPPTAAPTTAAP-----------------APNTTVT-- 632
          ||  ||: | | :   ||| : | | ||  | |                 ||   ||
Db    462 APAEAPSAAAPIVSTPPTTASVPETTAPPAAVPTEPIDVSVLSEAAIETPVAPPVEVTTE 521

Qy    632 ------VPPTAA-----------------------------------PTTAAPAPNTT 648
                ||  ||                                   | |: | ||
Db    522 VAVADVAPPEAAADLIIEPVEPPAPIPDLLEQTTSVPAVEAAESTSSPIPETSLPPPNEA 581

Qy    649 VTAPPTA-APTTA---APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPV--TTTSAPAT 702
          | :|  | || ||    | |  ::  |    |  ||  :  | :    ||  |:|| | |
Db    582 VASPEVAVAPITAPEPIPEPEPSLATPTEPIPVEAPVVIQEAVDAVEVPVTETSTSIPET 641

Qy    703 TPE------DDDIDPPLPNDPI---NPCVEECN 726
          | |       :  :|| :   |:    | |   |
Db    642 TVEFPEAVAEKVLDPAITEAPVTTQEPDVANIN 674


RESULT   4
VGBEX1
glycoprotein X precursor - equine herpesvirus 1 (strain Ab4p)
C;Species: equine herpesvirus 1
A;Note: host Equus caballus (domestic horse)
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 05-Sep-1997
C;Accession: H36802
R;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
submitted to GenBank, March 1992
A;Description: The DNA sequence of equine herpesvirus-1.
A;Reference number: A36805
A;Accession: H36802
A;Molecule type: DNA
A;Residues: 1-797 <TEL>
A;Cross-references: GB:M86664; NID:g330791; PID:g330862
R;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
Virology 189, 304-316, 1992
A;Title: The DNA sequence of equine herpesvirus-1.
A;Reference number: A41831; MUID:92295566
A;Contents: annotation; possible protein-coding frames
A;Note: neither amino acid nor nucleotide sequence is given
C;Genetics:
A;Gene: 71
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein h
C;Keywords: glycoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-797/Product: glycoprotein X #status predicted <MAT>
F;23-465/Region: serine/threonine-rich
F;489-797/Domain: equine herpesvirus 1 glycoprotein homology <EHG>
F;766-790/Domain: transmembrane #status predicted <TMN>
F;590/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             10.2%;  Score 446.5;  DB 1;  Length 797;
  Best Local Similarity   25.3%;  Pred. No. 1.6e-15;
  Matches  182;  Conservative   41;  Mismatches  186;  Indels  311;  Gaps   20;

Qy     78 SAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTT--TTQA 135
          | |:   || :  : |   : ||  | :|   |||  || :|  ||||| |||| || |
Db     77 STQSSSTAATSS-SAPSTASSTTSIPTSTSTETTTTTPTASTTTPTTTTAAPTTAATTTA 135

Qy    136 PTT---TQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTT-------------------TT 172
          ||    | | |||   | |  |||||   |||| | ||                    ||
Db    136 VTTAASTSAETTTATATATSTPTTTTPTSTTTTTATTTVPTTASTTTDTTTAATTTAATT 195

Qy    173 QAPTTTQAPT---TTQAPTTTQAPTT------ITQAATTPAATTPAATTPAATTPAATTP 223
          | ||| | |   || | ||| | ||       |.|||| |||| |||| |||| ||||
Db    196 TAATTTAATTTAATTTAATTTAATTTAATTSSATTAATTTAATTTAATTTAATTTAATTT 255
```

```
Qy    224 AATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCP 283
          |||| |  :|||                                                
Db    256 AATTTG--SPTS---------------------------------------------------- 266

Qy    284 EGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLP 343

Db    266 ---------------------------------------------------------------- 266

Qy    344 HGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACT 403
                                                  | |   | |              
Db    266 ----------------------------------------GSTSTTGASTSTP------- 279

Qy    404 GWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPAT 463
                                                                  | :|   
Db    279 ------------------------------------------------------SAST 282

Qy    464 QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPA- 523
           :  :|                 |||:   || |: ||  ||:| |:| ||   | : |   
Db    283 ATSATP---------------TSTSTSAAATTSTPT--PTSAATSAESTTEAPTSTPTT 324

Qy    523 -TAAPVPPTTAIPTPAPT---AAPTAAPTTA-------APESPT--TVTVPPTAAPTAAP 569
           |  |   |||  :|  |   |: |:| |||       :|:| |  | | |  |::  |  |  
Db    325 DTTTPSEATTATTSPESTTVSASTTSATTTAFTTESHTSPDSSTGSTSTAEPSSTFTLTP 384

Qy    570 TTAVPEI-PITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNT 628
          :|| |     | :|| | : :   :  ||    :      ||  |    : |  :       
Db    385 STATPSTDQFTGSSASTESDSTDSSTVPTTGTESITESSSTTEASTNLGSSTYESTEALE 444

Qy    629 TVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTA-------------- 675
          |     |:  || :|:| |   |     |       :||:    ||                
Db    445 TPDGNTTSGNTTPSPSPRTPSFADTQQTPDNGVSTQHTTINDHTTANAQKHAGHHRGRAG 504

Qy    675 APTAAPPTVAHAPNTTAAPVTTTSAPATTPEDD----------DIDPPLPNDPINPCVE 723
            :|   :|   ||  |   | :|  | :||           :| | ||  |  | :|    
Db    505 GRRGSPQGGSH---TTPHPDRLTPSPDDTYDDDTNHPNGRNNSIEIVPQLP--PDRPIIE 559


RESULT   5
A45155
mucin FIM-C.1 - African clawed frog (fragment)
C;Species: Xenopus laevis (African clawed frog)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Sep-1997
C;Accession: A45155
R;Hauser, F.; Hoffmann, W.
J. Biol. Chem. 267, 24620-24624, 1992
A;Title: P-domains as shuffled cysteine-rich modules in integumentary mucin C.1 (FIM-C.1
A;Reference number: A45155
A;Accession: A45155
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-662 <HAU>
A;Cross-references: GB:L02115; NID:g214147; PID:g951460
C;Superfamily: trefoil homology
F;162-202/Domain: trefoil homology <TRF1>
F;307-347/Domain: trefoil homology <TRF2>
F;354-394/Domain: trefoil homology <TRF3>
F;526-566/Domain: trefoil homology <TRF4>
F;573-613/Domain: trefoil homology <TRF5>
F;621-661/Domain: trefoil homology <TRF6>


  Query Match             9.8%;  Score 430.5;  DB 2;  Length 662;
  Best Local Similarity   22.4%;  Pred. No. 8.5e-15;
  Matches  194;  Conservative   27;  Mismatches  226;  Indels  419;  Gaps   28;

Qy     65 APRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTT----QAPTT--- 118
          ||   | |     :| |    | ||     |   | :|||| ||  |    :|||| 
Db     73 APATTAAGKAPTTAAATAPTTAAAG----APTTATGKAPATAAAPVPTTAASKAPTTAAA 128

Qy    118 ------TTQAPTTTTQA---------------------------------------- 129
                |||| |
Db    129 ATHSTAAAAAPTTAASAAKSKERSTSSSSEEEHCHVKPSKREMCGSKGITKKQCKKKNCC 188
```

```
Qy    129 ------------------------PTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTT 164
                                     ||||:|||| |  |||  ||||   |  |  |||  
Db    189 FDPKGHGGIHCFHRKPKGHSHEEHTTTTTKAPTTIQIATTTTTPTTTTTTTKATPTTTTT 248

Qy    165 TQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPA 224
          |:|  |||   |||:| |||  ||||   || |:| |||  ||   |||  ||  |||   
Db    249 TKATPTTT---TTTKATTTTTTPTTT---TTTTKATTTPTTTT--TTTPTTTTTKATTTT 300

Qy    225 ATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPE 284
           || |                                                          
Db    301 TTTSG---------------------------------------------------- 306

Qy    285 GLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPH 344
                        ||  | |                                        
Db    306 ----------------ECKMEPS---------------------------------- 313

Qy    345 GNRCDKYYQCVHGNLVERRCGA-GTHFSFELQQ---CDHIELVGCTLPGGESEEVDVDED 400
               |  : : | :|   |  |   : |   | :      ||    |:  |        
Db    313 -----KREDCGYSGITESQCRTKGCCFDSSIPQTKWCFY------TL----SQVAD---- 354

Qy    401 ACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFS 460
              | || :                 |||                              
Db    354 ------CKVEPSQ--------------------RVDCG---------------------- 366

Qy    461 PATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATP 520
                   |    :|   | |  ::      :       : :  ||  |   | |  
Db    366 ----------FRGITADQCRQKNCCFDSSI-----SGTKWCFYSTSQVAATKTTTPTT- 410

Qy    521 PATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITV 580
                |||   | | |  ||   ||     ||| |                        
Db    410 -----TTTPTTTTTTKATTTTPTTTTTT-----PTTTTT---------------------- 439

Qy    581 TSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTT 640
                |   | | |  |||  |   || |   |  |||    | || |   |  |||  
Db    439 --------TTTTTKATTTTPTTTTP--TTTTTKATTTTPTTTTTTPTTTTTKATTTTPTT 488

Qy    641 AAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAP 700
             | || |   |  |||      || |                    || |  ||||  
Db    489 TTTPTTTTTTKATTTTPTT------TTTTT----------------TTTKATTTTTSGE 525

Qy    701 ATTPEDDDIDPPLPNDPINPC----------------------VEECNVLPWAHADC-- 736
                |   |    : |                        | :| | | :  ||    
Db    526 CKMEPSKRADCGYPGITESQCRSKGCCFDSSIPQTKWCFYSLPQVADCKVAPSSRVDCGF 585

Qy    736 --------------------DKYWVCDGNNQVLVVCSEGLQFNPTTKTCD---------F 766
                              |    :|   :||      | ||  |               
Db    586 GGITADQCRQRNCCFDSSISGTKWCFYSTSQGNAMCS-----GPPTKRRDCGYPGISSSV 640

Qy    767 ACNVGCVRSNIQMSESYEGVQVFIPW 792
            | ||   |  |:         :||  
Db    641 CINRGCCWDNSVMN---------VPW 657


RESULT   6
A35887
mucin precursor, pancreatic - human
C;Species: Homo sapiens (man)
C;Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 03-Feb-1994
C;Accession: A35887
R;Lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A;Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A;Reference number: A35887; MUID:90368716
A;Accession: A35887
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1335 <LAN>
A;Cross-references: GB:J05582
C;Superfamily: polymorphic epithelial mucin


  Query Match             9.7%;  Score 428;  DB 2;  Length 1335;
```

```
  Best Local Similarity   26.7%;  Pred. No. 2e-14;
  Matches  208;  Conservative   45;  Mismatches  294;  Indels  232;  Gaps   37;

Qy      17 RPEVSDAEKNPALHE--PHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRDCAPGTE 74
           || :      |  |:    ||  ||      |           :|:       |||:
Db     105 RPALGST--TPPAHDVTSAPDNKPAPGSTAPPA----------HGVTSAPDTRPAPGS- 151

Qy      75 FKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTT-------TQAPTTTTQ 127
             :|       |  | | | | :    |   | || |  || :|       | || |
Db     151 ---TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 207

Qy     128 APTTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTTQAPTTTTQAPTTTQAPTTTQA 186
           :|   |   | || |  || :|  |    | || |     :|   |  | || |  |
Db     208 PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 267

Qy     187 PTTTQAPT-TITQAATT---PAATTPAA---TTPAATTPAATTPAATTPGVPAPTSAP-- 238
           | :|  |   :| |  |   | :| | |   |:   | ||   | :| |    ||||
Db     268 PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA---PGSTAPPAHGVTSAPDT 324

Qy     238 ------VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTF--EQRCPEGLYFN 289
                 ||                                ::| |   : |   |
Db     325 RPAPGSTAPP------------------------------AHGVTSAPDTRPAPGSTAP 353

Qy     290 PYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCD 349
           |      :|      |  | |||
Db     354 PAHGVTSAPDTRPAPG--STAPP----------------------------------- 375

Qy     350 KYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP- 409
             ||                |  | | |          |     |   |
Db     375 -----AHG--------------VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 414

Qy     409 ----TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCPGNLHFSP 461
               |   : | |                  |      || | |   || ||:    |
Db     415 AHGVTSAPDTRPAP------------------GSTAPPAHGVTSAPDTRPAPGST-APP 454

Qy     462 ATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVV--PPAT 519
           |     :|          |:    |||  |    |:||    |    || |:  |   | |
Db     455 AHGVTSAP----------DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 504

Qy     520 PPATAAPVPP---TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEI 576
            ||  :  ||    |: |   |    || |      :| |   | : || |   |: |:
Db     505 RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 564

Qy     577 -PITVTSAP-----TAAPTAAPTAAPTAAPTTAVPEIPTT-----VTSPP----TAAPTT 621
            |   ::||     |:||   |    || |   |   | |       |:||    |:|| |
Db     565 RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 624

Qy     622 AAPAPNTTVTVPP----TAAPTTAAPAPNTTVTAPP----TAAPTTAAPAPNTTVTVPPT 673
             ||| :  | ||    |:|| |  ||| :  ||||    |:|| |  ||| :  | ||
Db     625 -RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPA 675

Qy     674 AAPTAAPPT------------VAHAPNTTAAPVTT-------TSAPATTPEDDDIDPP 712
           |:|| |              |  ||:|  || :|       |||| | |      ||
Db     676 HGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 734


RESULT   7
B35175
episialin B - human
C;Species: Homo sapiens (man)
C;Date: 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change 23-Feb-1997
C;Accession: B35175
R;Ligtenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A;Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene enc
A;Reference number: A35175; MUID:90202794
A;Accession: B35175
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1255 <LIG>
A;Cross-references: GB:M34088; GB:J05288
C;Superfamily: polymorphic epithelial mucin
C;Keywords: cytoskeleton; transmembrane protein


  Query Match             9.7%;  Score 428;  DB 2;  Length 1255;
  Best Local Similarity   26.7%;  Pred. No. 1.9e-14;
  Matches  208;  Conservative   45;  Mismatches  294;  Indels  232;  Gaps   37;

Qy      17 RPEVSDAEKNPALHE--PHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRDCAPGTE 74
           || :      |  |:    ||  ||      |           :|:       |||:
Db     105 RPALGST--TPPAHDVTSAPDNKPAPGSTAPPA----------HGVTSAPDTRPAPGS- 151

Qy      75 FKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTT-------TQAPTTTTQ 127
             :|       |  | | | | :    |   | || |  || :|       | || |
Db     151 ---TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 207

Qy     128 APTTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTTQAPTTTTQAPTTTQAPTTTQA 186
           :|   |   | || |  || :|  |    | || |     :|   |  | || |  |
Db     208 PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 267

Qy     187 PTTTQAPT-TITQAATT---PAATTPAA---TTPAATTPAATTPAATTPGVPAPTSAP-- 238
           | :|  |   :| |  |   | :| | |   |:   | ||   | :| |    ||||
Db     268 PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA---PGSTAPPAHGVTSAPDT 324

Qy     238 ------VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTF--EQRCPEGLYFN 289
                 ||                                ::| |   : |   |
Db     325 RPAPGSTAPP------------------------------AHGVTSAPDTRPAPGSTAP 353

Qy     290 PYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCD 349
           |      :|      |  | |||
Db     354 PAHGVTSAPDTRPAPG--STAPP----------------------------------- 375

Qy     350 KYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP- 409
             ||                |  | | |          |     |   |
Db     375 -----AHG--------------VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 414

Qy     409 ----TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCPGNLHFSP 461
               |   : | |                  |      || | |   || ||:    |
Db     415 AHGVTSAPDTRPAP------------------GSTAPPAHGVTSAPDTRPAPGST-APP 454

Qy     462 ATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVV--PPAT 519
           |     :|          |:    |||  |    |:||    |    || |:  |   | |
Db     455 AHGVTSAP----------DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 504

Qy     520 PPATAAPVPP---TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEI 576
            ||  :  ||    |: |   |    || |      :| |   | : || |   |: |:
Db     505 RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 564

Qy     577 -PITVTSAP-----TAAPTAAPTAAPTAAPTTAVPEIPTT-----VTSPP----TAAPTT 621
            |   ::||     |:||   |    || |   |   | |       |:||    |:|| |
Db     565 RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 624

Qy     622 AAPAPNTTVTVPP----TAAPTTAAPAPNTTVTAPP----TAAPTTAAPAPNTTVTVPPT 673
             ||| :  | ||    |:|| |  ||| :  ||||    |:|| |  ||| :  | ||
Db     625 -RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPA 675

Qy     674 AAPTAAPPT------------VAHAPNTTAAPVTT-------TSAPATTPEDDDIDPP 712
           |:|| |              |  ||:|  || :|       |||| | |      ||
Db     676 HGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 734


RESULT   8
A35886
polymorphic epithelial mucin precursor - human
C;Species: Homo sapiens (man)
C;Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 03-Feb-1994
C;Accession: A35886
R;Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burc
J. Biol. Chem. 265, 15286-15293, 1990
A;Title: Molecular cloning and expression of human tumor-associated polymorphic epith
A;Reference number: A35886; MUID:90368715
A;Accession: A35886
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1295 <GEN>
```

```
A;Cross-references: GB:J05581
C;Superfamily: polymorphic epithelial mucin

  Query Match              9.7%;  Score 428;  DB 2;  Length 1295;
  Best Local Similarity   26.7%;  Pred. No. 2e-14;
  Matches  208;  Conservative   45;  Mismatches  294;  Indels  232;  Gaps   37;

Qy      17 RPEVSDAEKNPALHE--PHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRDCAPGTE 74
           || :       |  |:    ||  ||       |            :|:        |||:
Db     105 RPALGST--TPPAHDVTSAPDNKPAPGSTAPPA----------HGVTSAPDTRPAPGS- 151

Qy      75 FKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTT-------TQAPTTTTQ 127
             :|        |  | | | :    |   | || |  || :|       | || |
Db     151 ---TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 207

Qy     128 APTTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTTQAPTTTTQAPTTTQAPTTTQA 186
             :|   |   | || |  || :|  |    | || |     :|   |   | || |  |
Db     208 PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 267

Qy     187 PTTTQAPT-TITQAATT---PAATTPAA---TTPAATTPAATTPAATTPGVPAPTSAP-- 238
           | :|  |   :| |  |   | :| | |   |:   | ||   | :| | |     ||||
Db     268 PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA---PGSTAPPAHGVTSAPDT 324

Qy     238 ------VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTF--EQRCPEGLYFN 289
                 ||                          ::| |   : |   |
Db     325 RPAPGSTAPP----------------------------AHGVTSAPDTRPAPGSTAP 353

Qy     290 PYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCD 349
           |      :|      |  | |||
Db     354 PAHGVTSAPDTRPAPG--STAPP------------------------------------ 375

Qy     350 KYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP- 409
                 ||                    |  | |          |     |   |
Db     375 -----AHG--------------VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 414

Qy     409 ----TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCPGNLHFSP 461
               |   :  | |                |     || | |   || ||:    |
Db     415 AHGVTSAPDTRPAP------------------GSTAPPAHGVTSAPDTRPAPGST-APP 454

Qy     462 ATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVV--PPAT 519
           |     :|          |:     ||| |    |:||   |    || |:  |   | |
Db     455 AHGVTSAP----------DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 504

Qy     520 PPATAAPVPP---TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEI 576
            ||  :  ||    |: |   |    || |       :| |   | : || |   |: |:
Db     505 RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 564

Qy     577 -PITVTSAP-----TAAPTAAPTAAPTAAPTTAVPEIPTT-----VTSPP----TAAPTT 621
            |   ::||     |:||   |    || |   |   | |      |:||    |:|| |
Db     565 RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 624

Qy     622 AAPAPNTTVTVPP----TAAPTTAAPAPNTTVTAPP----TAAPTTAAPAPNTTVTVPPT 673
             ||| :  | ||    |:|| |  ||| :  ||||    |:|| |  ||| :  | ||
Db     625 -RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPA 675

Qy     674 AAPTAAPPT------------VAHAPNTTAAPVTT-------TSAPATTPEDDDIDPP 712
             |:|| |              |  ||:|  || :|       |||| |  |      ||
Db     676 HGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 734


RESULT   9
A35175
episialin A - human
C;Species: Homo sapiens (man)
C;Date: 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change 23-Feb-1997
C;Accession: A35175
R;Ligtenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A;Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene enc
A;Reference number: A35175; MUID:90202794
A;Accession: A35175
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1264 <LIG>
A;Cross-references: GB:M32738; GB:M34088; GB:J05288
C;Superfamily: polymorphic epithelial mucin
C;Keywords: cytoskeleton; transmembrane protein

  Query Match              9.7%;  Score 428;  DB 2;  Length 1264;
  Best Local Similarity   26.7%;  Pred. No. 1.9e-14;
  Matches  208;  Conservative   45;  Mismatches  294;  Indels  232;  Gaps   37;

Qy      17 RPEVSDAEKNPALHE--PHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRDCAPGTE 74
           || :       |  |:    ||  ||       |            :|:        |||:
Db     114 RPALGST--TPPAHDVTSAPDNKPAPGSTAPPA----------HGVTSAPDTRPAPGS- 160

Qy      75 FKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTT-------TQAPTTTTQ 127
             :|        |  | | | :    |   | || |  || :|       | || |
Db     160 ---TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 216

Qy     128 APTTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTTQAPTTTTQAPTTTQAPTTTQA 186
             :|   |   | || |  || :|  |    | || |     :|   |   | || |  |
Db     217 PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 276

Qy     187 PTTTQAPT-TITQAATT---PAATTPAA---TTPAATTPAATTPAATTPGVPAPTSAP-- 238
           | :|  |   :| |  |   | :| | |   |:   | ||   | :| | |     ||||
Db     277 PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA---PGSTAPPAHGVTSAPDT 333

Qy     238 ------VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTF--EQRCPEGLYFN 289
                 ||                          ::| |   : |   |
Db     334 RPAPGSTAPP----------------------------AHGVTSAPDTRPAPGSTAP 362

Qy     290 PYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCD 349
           |      :|      |  | |||
Db     363 PAHGVTSAPDTRPAPG--STAPP------------------------------------ 384

Qy     350 KYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP- 409
                 ||                    |  | |          |     |   |
Db     384 -----AHG--------------VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 423

Qy     409 ----TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCPGNLHFSP 461
               |   :  | |                |     || | |   || ||:    |
Db     424 AHGVTSAPDTRPAP------------------GSTAPPAHGVTSAPDTRPAPGST-APP 463

Qy     462 ATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPSTVV--PPAT 519
           |     :|          |:     ||| |    |:||   |    || |:  |   | |
Db     464 AHGVTSAP----------DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 513

Qy     520 PPATAAPVPP---TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEI 576
            ||  :  ||    |: |   |    || |       :| |   | : || |   |: |:
Db     514 RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 573

Qy     577 -PITVTSAP-----TAAPTAAPTAAPTTAVPEIPTT-----VTSPP----TAAPTT 621
            |   ::||     |:||   |    || |   |   | |      |:||    |:|| |
Db     574 RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 633

Qy     622 AAPAPNTTVTVPP----TAAPTTAAPAPNTTVTAPP----TAAPTTAAPAPNTTVTVPPT 673
             ||| :  | ||    |:|| |  ||| :  ||||    |:|| |  ||| :  | ||
Db     634 -RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPA 684

Qy     674 AAPTAAPPT------------VAHAPNTTAAPVTT-------TSAPATTPEDDDIDPP 712
             |:|| |              |  ||:|  || :|       |||| |  |      ||
Db     685 HGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 743


RESULT  10
S49915
extensin-like protein - maize
C;Species: Zea mays (maize)
C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 09-Sep-1997
C;Accession: S49915
R;Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A;Description: Pex genes: pollen-specific genes with extensin-like domains.
```

A;Reference number: S49915
A;Accession: S49915
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1188 <RUB>
A;Cross-references: EMBL:Z34465; NID:g600117; PID:g600118

Query Match 9.6%; Score 421.5; DB 2; Length 1188;
Best Local Similarity 22.7%; Pred. No. 3.8e-14;
Matches 175; Conservative 89; Mismatches 323; Indels 183; Gaps 29;

```
Qy       2 IKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDX-----PPAEQXXLLPXEYDCTKFYY 56
Db     520 VKTTSPPAPIGSPSPPPPVSVVSPPPPVKSPPPPAPVGSPPPPEKSPPPP---------- 570

Qy      57 CEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAP------- 110
Db     570 --------APVASPPPPVKSPPPPTLVASPPPPVKSPPPPAPVASPPPPVKSPPPPTPVA 621

Qy     110 -TTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAP 168
Db     622 SPPPPAPVASSPPPMKSPPPPTPVSSPPPPEKSPPPPPPAKSTPPPEEYPTPPTSVKSSP 681

Qy     169 TTTTQAPTTTQAP------------TTTQAPTTTQAPTTITQAATTPAATTPAATTPA-- 215
Db     682 PPEKSLPPPTLIPSPPPQEKPTPPSTPSKPPSSPEKPSPPKEPVSSPPQTPKSSPPPAPV 741

Qy     215 -------ATTPAATTPAATTPGV-----PAPTSAPVWPPICELLPNGCPADFDIHLLIPH 262
Db     742 SSPPPTPVSSPPALAPVSSPPSVKSSPPPAPLSSP--PPAPQV--KSSPPPVQVSSPPPA 797

Qy     263 DKYCNLFYQCSNGYTFEQRC-PEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDED 321
Db     798 PKSSPPLAPVSSPPQVEKTSPPPAPLSSPPLAPKSSPPHV----VVSSPPPVVKSSPP-- 852

Qy     322 IDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIE 381
Db     852 -----------PAPVSSPPLTP-----------------KPASPPAHVS------SPPE 876

Qy     382 LVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYL 441
Db     877 VVKPSTPPAPTTVISP----------PSEPKSSPP-----PTPVSLPPPIVKSS------ 916

Qy     442 QCVHGQTIARPCPGNLHFSPAT-QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPT 500
Db     916 ----------PPPAMVSSPPMTPKSSPPPVVV--------SSPPPTVKSSPPPAPVSSPP 957

Qy     501 AAPTAAPTAAPSTVVPP---ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTV 557
Db     958 ATPKSSPPPAPVNLPPPEVKSSPPPTPVSSPP------PAPKSSPPPAPMSSPP--PPEV 1009

Qy     558 TVPPTAAPTAAPTTAV----PEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTS 613
Db    1010 KSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSP---PPPAPVSSP-------PPPVKS 1059

Qy     614 PPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPT 673
Db    1060 PPPPAPISSPPPP---VKSPPPPAPVSSPPPP---VKSPPPPAPVSSPPPP---IKSPPP 1110

Qy     674 AAPTAAPPTVAHAPNTTAAPVTTTS-APATTP-----EDDDIDPPLPNDP 717
Db    1111 PAPVSSPPPAPVKPPSLPPPAPVSSPPPVVTPAPPKKEEQSLPPPAESQP 1160
```

RESULT 11
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR019c
C;Species: Saccharomyces cerevisiae
C;Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 05-Dec-1997
C;Accession: S48478; A26877; B26877; S27281; JC6123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: GB:Z47047; EMBL:Z38061; MIPS:YIR019c; NID:g603997; PID:g763364
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>
A;Cross-references: EMBL:M16164; NID:g172522; PID:g172525
A;Accession: B26877
A;Molecule type: DNA
A;Residues: 762-1331 <YA2>
A;Cross-references: EMBL:M16165; NID:g172523; PID:g172526
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacc
A;Reference number: S27281
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseud
A;Reference number: JC6123
A;Accession: JC6123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: GB:U30626; NID:g1304386; PID:g1304387
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2
A;Cross-references: MIPS:YIR019c; SGD:S0001458
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F;5-21/Domain: transmembrane #status predicted <TM1>
F;1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 9.3%; Score 410; DB 2; Length 1367;
Best Local Similarity 19.9%; Pred. No. 1.6e-13;
Matches 152; Conservative 124; Mismatches 238; Indels 248; Gaps 28;

```
Qy      32 PHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCT 91
Db     288 PHHDTTP-----------CTK-------KKTTTSKTCTKKTTTPVPTPSSSTTESSSAP 328

Qy      92 LPGPPAETTQ---APATTQAPTTTQAPTTTTQAPTT------TTQAPTTTTQAPTTTQAP 142
Db     329 VPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTT 388

Qy     143 TTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTT------TQAPTTTQAPTTTQAPTT- 196
Db     389 ESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTS 448

Qy     196 -ITQAATTPAATTPAATTPAATTP--AATTPAATTPGVPAP-------TSAPVWPPICEL 245
Db     449 STTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAP-VPTPSSSTTESSSAPV------- 501

Qy     246 LPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDG 305
Db     501 -------------------------TSSTTESSSAP----------VPTPSSSTTES 522

Qy     306 EISPAP-PVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRC 364
Db     523 SSAPAPTPSSSTTESSSAPV---------------------------------------- 543
```

```
Qy    365 GAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPAD 424
Db    543 --------------------TSSTTESSSAPV-----------------PTPSSSTTE 563

Qy    425 FSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQ 484
Db    564 SS----------------------STPVT-----SSTTESSSAPVPT-------PSSST 588

Qy    485 CTSTAAPTAAP-------TAAPTAAPTAAPTAAPSTVVPPATPPATAAPVP-------PT 530
Db    589 TESSSAPVPTPSSSTTESSSAPAPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTES 648

Qy    531 TAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPIT-VTSAPTAAPT 589
Db    649 SSAPVPTPSSSTTESSSAPVP-TPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPV 707

Qy    590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVT----VPPTAAPTTAAPAP 645
Db    708 PTPSSSTTESSSAPVP-TPSSSTTESSSAP---VPTPSSSTTESSSAPVTSSTTESSSAP 763

Qy    646 NTTVTAPPTAAPTTAAPAPNTTVT----VP-PTAAPTAAPPTVAHAPNTTAAPVTTTSAP 700
Db    764 VPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSVAPVPTPSSSSNITSSAP 823

Qy    701 ATTP-----EDDDIDPPLPND--------PINPCVEECNVLP 729
Db    824 SSTPFSSSTESSSVPVPTPSSSTTESSSAPVSSSTTESSVAP 865
```

RESULT 12
A37232
mucin, tracheal (AMN-22) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 30-Sep-1993
C;Accession: A37232
R;Gerard, C.; Eddy Jr., R.L.; Shows, T.B.
J. Clin. Invest. 86, 1921-1927, 1990
A;Title: The core polypeptide of cystic fibrosis tracheal mucin contains a tandem repeat
A;Reference number: A37232; MUID:91072667
A;Accession: A37232
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-294 <GER>
A;Cross-references: GB:M57417

Query Match 8.7%; Score 384; DB 2; Length 294;
Best Local Similarity 24.1%; Pred. No. 8.4e-13;
Matches 141; Conservative 8; Mismatches 132; Indels 304; Gaps 15;

```
Qy    100 TQAPATTQAPTTTQAPTTTTQ---APT-TTTQAP-----TTTTQAPTTTQAPTTTQAPTT 150
Db      3 TPTPTGTQTPTPTPITTTTTMVTPTPTIITSTQTPTPTPITTTTVTPTPT--PTSTQRTTP 60

Qy    151 TQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAAT-TPAATTPA 209
Db     61 TSITTTTTVTPTPT---PITTTTTVTPTPTPTGTQTPTTTPISTTTTVTPTPTPTGTQTL 117

Qy    210 ATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLF 269
Db    118 TPTPITTT-TTVTPTPTPTGTQTPTSTPI---------------------------------- 146

Qy    270 YQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLD 329
Db    146 ---------------------------------------------------------- 146

Qy    330 NGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPG 389
Db    146 ---------------------------------------------------------- 146

Qy    390 GESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTI 449
Db    146 ---------------------------------------------------------- 146
```

```
Qy    450 ARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTA 509
Db    146 ----------------------------------------TTTTTVTPTPTPTGTQT- 163

Qy    510 APSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAP 569
Db    163 ------PTTTPITTTTTVTPT---PTPTGTRYPTPTPIT------TTTTVTPTPTPTSTQ 207

Qy    570 TTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPT-TVTSPPTAAPTTAAPAPNT 628
Db    208 STTPTPITTTNTVTPTPTPTGTQTPTPTPITTTTTMVTPTPTITSTQTPTPT-----PIT 262

Qy    629 TVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPT 673
Db    263 TTTVTPT-------PTPTSTQRTTPTSITTT------TTVTPTPT 294
```

RESULT 13
A38420
antifreeze glycoprotein precursor - black rockcod
C;Species: Notothenia coriiceps neglecta (black rockcod)
C;Date: 28-Jun-1991 #sequence_revision 30-Jan-1993 #text_change 09-Sep-1997
C;Accession: A38420
R;Hsiao, K.C.; Cheng, C.H.C.; Fernandes, I.E.; Detrich, H.W.; DeVries, A.L.
Proc. Natl. Acad. Sci. U.S.A. 87, 9265-9269, 1990
A;Title: An antifreeze glycopeptide gene from the antarctic cod Notothenia coriiceps
A;Reference number: A38420; MUID:91067687
A;Accession: A38420
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-822 <HSI>
A;Cross-references: GB:M55000; NID:g213391; PID:g213392
A;Note: the authors translated the codon CCT for residue 489 as Ala
C;Keywords: glycoprotein

Query Match 8.5%; Score 374; DB 2; Length 822;
Best Local Similarity 25.7%; Pred. No. 6e-12;
Matches 193; Conservative 21; Mismatches 332; Indels 204; Gaps 26;

```
Qy     65 APRDCAPGTEFKFSAQTCVHAA---------LAGCTLPGPP----AETTQAPATTQAPTT 111
Db    146 AATPATPAFHFAATAATPATAATPALIFAATAATAATPATPAFHFAATAATPATAATPAL 205

Qy    112 TQAPTTTTQA-PTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTT 170
Db    206 IFAATAATAATPATAALNFAATAATPATAATPALIFAATAATAATPATAALNFAATAATA 265

Qy    171 TTQAPTTTQAPTTTQAPTTTQAPTTITQA----ATTPA---------ATTPA-ATTP--- 214
Db    266 ATPATAACNFAATAATPATAATPALIFAATAATAATPATAACNFAATAATPATAATPALI 325

Qy    214 -AATTPAATTPA------ATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYC 266
Db    326 FAATAATAATPATAALNFAATAATPATAATP---------------------------- 357

Qy    267 NLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGD 326
Db    357 -----------------ALIFAATAATAATPATAALNFAATAATPATAAT-PALIFAAT 397

Qy    327 LLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGT----HFSFELQQCDHIEL 382
Db    398 AATAATPATAAL------------------NFAATAATAATPATPAFNF--------AA 430

Qy    383 VGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQ 442
Db    431 TAATPATAATPALIFAATAATA-----------ATPATAALNFAATAATPATAATPALIF 479

Qy    443 CVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAA 502
Db    480 AATAATAATPATAALHFAATAATAATPATAAL--------NFAATAATPATAATPALIFA 531
```

```
Qy    503 PTAAPTAAPST-------VVPPATPPATA-------APVPPTTAIPTP--APTAAPTAAP 546
Db    532 ATAATAATPATAAFNFAATAATAATPATAALNFAATAATPATAATPALIFAATAATAATP 591

Qy    547 TTAAPESPTTVTVPPTAAPTA------APTTAVPEIP---------------------- 578
Db    592 ATAALNFAATAATPATAATPALIFAATAATAATPATPAFHFAATAATPATAATPALIFAA 651

Qy    578 --ITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAA-PTTAAPAPNTTVTVPP 634
Db    652 TAATAATPATAALNFAATAATAATPATAALNFAATAATPATAATPALIFAATAATAATPA 711

Qy    635 TAA---------PTTAA-PAPNTTVTAPPTAAPTTAA---PAPNTTVTVPPTAAPT---- 678
Db    712 TAALNFAATAATPATAATPALIFAATAATAATPATAAFNFAATAATAATPATAATPALIF 771

Qy    678 AAPPTVAHAPNTTAAPV---TTTSAPATTP 704
Db    772 AATAATAATPATPATPALIFAATAATAATP 801


RESULT  14
A48292
mucin, tracheobronchial - dog
C;Species: Canis lupus familiaris (dog)
C;Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 10-Sep-1997
C;Accession: A48292
R;Verma, M.; Davidson, E.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 7144-7148, 1993
A;Title: Molecular cloning and sequencing of a canine tracheobronchial mucin cDNA contai
A;Reference number: A48292
A;Accession: A48292
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1118 <VER>
A;Cross-references: GB:X69164; NID:g402557; PID:g402558
A;Note: authors translated the codon CAA for residue 52 as Gly, GAT for residue 94 as As
C;Comment: It is uncertain whether Met-1 or Met-52 is the initiator.
C;Keywords: glycoprotein


  Query Match             8.5%;  Score 373;  DB 2;  Length 1118;
  Best Local Similarity   23.5%;  Pred. No. 8.7e-12;
  Matches  209;  Conservative   52;  Mismatches  309;  Indels  320;  Gaps   37;

Qy     66 PRDCAPGTEFKFSAQTCVHAALAGCTLP----------GPPAETTQAPATTQAPTTTQAP 115
Db    203 PPQCLTPTPTPLSTPSIIRPQPEALTLPLCLSAVSETTPTPISTTTTVTPTPTPTGTQTP 262

Qy    116 TTTTQAPTTTTQAPTTT-TQAPTT---------------------------------TQ 140
Db    263 TTTPITTTTVTPTPTPTGTQTPTTVLITTTTTMPTPTPTSQKLRGSDTPVLVIALVTPTP 322

Qy    141 APTTTQAPTTTQAPTTTTQAPTTT---TQAPTTTTQAPTTTQAPT-------------- 183
Db    323 TPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTIRITTTTTMTPTPTTTSTKVLVTPYTT 382

Qy    183 ------TTQAPTTTQAPTTITQAATTPAATTPAATTPA----------------ATTPAA 220
Db    383 TTGVTPTTHRHTRHQTTTTHTTYTTVTAIATPTGTRPTTYLHTTTTVTQTTHRQTTRPQP 442

Qy    221 TTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQ 280
Db    443 TTPSPTTTTVD-PTQHPSPPPLGDPNPN-THRDTDPN----HDTTTTT--TVTPTPTPTR 494

Qy    281 RCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNED----------------EDI 322
Db    495 TRPQPRYYTTTTTMTPTP---------TPPTPHTDHHHDPLPPPPPQVRHHRTPRHCHEL 545

Qy    323 DIGDLL---DNGCPANFEIDWLLPHGNRCDKYYQCVH-----------GNLVERRCGAG 367
Db    546 TSGSPLTHYPNTPTTN------LLEKNRSELINTPTHTMIISSAAMATPPSLVHSRCRSG 599

Qy    368 THFSFELQQ-----CDHIEL-------VGCTLPGG------------ESEEVDVDEDAC 402
Db    600 -HSSKAQRKSRASGCGLMMLGVESGVGVGSTHAGGALVRSRCRFTRHSRAEE-------- 651

Qy    403 TGWYCPTEPIEWEPLPNGCPADFSIDHLLPHES-------DCGQYLQCVHGQTIARPCPG 455
Db    651 ----------------GLGPEAVVSMMLGVESGVGVGVRHCGGRLHSAGGG--AEDDPG 691

Qy    456 NL-----------HFSPATQSCESPV-----------TAGCQVFECDSDNQCTSTAAPTA 493
Db    692 RVPRVGALTDSTQHTSRKAIKCKAGVPNEELTTYIACASALQSETCAADMIATATIHTTT 751

Qy    494 APT-------AAPTAAPTAAPTAAPSTVV----------------PPATPPATAAPVPPT 530
Db    752 SRSIRKPKTISELTTYACASTAASVGTVVQLDMIATGGLRDRGQMAGAMPAGDSRGSPGT 811

Qy    531 TAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTA 590
Db    812 ELEFRPNGLQAPTPTPIS------TTTTVTPTPTPTG------------TQTPTTTPSP 852

Qy    591 APTAAPTAAPTT--AVPEIPTTVTSPPTAAPTT-------------------------- 622
Db    853 PPPGDPTKGKTVYKIIGSLRTTFNLPAWLGYITICKCLMQSNRKTKTAETINVKTLGSLM 912

Qy    622 --AAPAPNTTVTVP-PTAAPTTAAPAPNTT--------------VTAPPT--AAPTTAA 661
Db    913 MDVTPINSCRVTPPFPSGKPCADSRLMNPTAETGSVIGAGCSCCVLTTPTTHRVIRCTTA 972

Qy    662 PAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDP 711
Db    973 PTP-TPISTTTTVTPTQHPP--AHRP--TTSPITTTTT-VTQPDTTHTDP 1016


RESULT  15
S15921
protein TPX-VT3 - Thermoproteus phage 1 (strain strain KRA1 10/12)
C;Species: Thermoproteus phage 1, TTV1
A;Variety: strain KRA1 10/12
C;Date: 21-Nov-1993 #sequence_revision 26-May-1995 #text_change 09-Sep-1997
C;Accession: S15921; S06686
R;Neumann, H.; Zillig, W.
Nucleic Acids Res. 18, 2171, 1990
A;Title: Nucleotide sequence of the viral protein TPX of the TTV1 variant VT3.
A;Reference number: S15921; MUID:90245666
A;Accession: S15921
A;Molecule type: DNA
A;Residues: 1-474 <NEU>
A;Cross-references: EMBL:X14717; NID:g62191; PID:e20711; PID:g584349
A;Experimental source: strain KRA1 10/12
A;Note: the authors translated the initiation codon GTG for residue 1 as Val
R;Neumann, H.
submitted to the EMBL Data Library, March 1989
A;Reference number: S06686
A;Accession: S06686
A;Molecule type: DNA
A;Residues: 207-474 <NEW>
A;Cross-references: EMBL:X14717
A;Experimental source: strain KRA1 10/12
C;Genetics:
A;Start codon: GTG


  Query Match             8.4%;  Score 367.5;  DB 2;  Length 474;
  Best Local Similarity   32.6%;  Pred. No. 8e-12;
  Matches  107;  Conservative   20;  Mismatches  132;  Indels   69;  Gaps   13;

Qy    394 EVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPC 453
Db    192 EIDISQFAAF-----TDPLVMYLSINGSTANVKV-----------YKQ---GSNI---- 228

Qy    454 PGNLHFSPATQSCESPVTAGCQ----------VFECDSDNQCTSTAAPTAAPTAAPTAAP 503
Db    228 -GTVSGNYSTTPYGNPSMAGYGTVDKHYANFIVLPYEPDPQVTVTPISSPSPTPTPTPTP 286
```

```
Qy     504 TAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTA 563
           |  ||  |:   |  ||  |  | |  |  ||| ||  ||  ||     :||     ||
Db     287 TPTPTPTPT---PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT-----PTP 338

Qy     564 APTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPT-TA 622
            ||  ||          |  ||  ||  ||  ||  ||  :  :   ||  ||  || |
Db     339 TPTPTPTP---------TPTPTPTPTPTPTPTPTPTPTYDITYVVFDVTPSPTPTPTPTP 389

Qy     623 APAPNTTVTVPPTAAPT-TAAPAPNTTVTAPPTAAPT-TAAPAPNTT---------VTVP 671
           | |  | |  ||  || |  | |  | |  ||  || |  | |  |          ||
Db     390 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTYDITYVIFDVTPS 449

Qy     672 PTAAPTAAPPTVAHAPNTTAAPVTTTSA 699
           ||  ||  |      |  |  | :|||:
Db     450 PTPTPTPTP-----TPTPTPTPTSTTSS 472
```

Search completed: July 18, 1999, 00:45:11
Job time: 267 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 01:41:31 ; Search time 19.54 Seconds
(without alignments)
1108.501 Million cell updates/sec

Title: US-09-294-663-4
Perfect score: 4395
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 807

Scoring table: BLOSUM62

Searched: 74019 seqs, 26840295 residues

Database : SwissProt_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	670	15.2	5179	1	MUC2_HUMAN	Q02817 homo sapien
2	457	10.4	865	1	CPN_DROME	Q02910 drosophila
3	446.5	10.2	797	1	VGLX_HSVEB	P28968 equine herp
4	430.5	9.8	662	1	MUC1_XENLA	Q05049 xenopus lae
5	428	9.7	1255	1	MUC1_HUMAN	P15941 h mucin 1 p
6	410	9.3	1367	1	AMYH_YEAST	P08640 saccharomyc
7	372.5	8.5	1664	1	SLP1_CLOTM	Q06852 clostridium
8	369.5	8.4	822	1	ANP_NOTCO	P24856 notothenia
9	367.5	8.4	474	1	VTP3_TTV1V	P19275 thermoprote
10	357	8.1	279	1	Y091_NPVOP	O10341 orgyia pseu
11	352	8.0	670	1	VG50_HSVI1	Q00130 ictalurid h
12	336.5	7.7	267	1	EXTN_MAIZE	P14918 zea mays (m
13	331	7.5	400	1	MUA1_XENLA	P10667 xenopus lae
14	321	7.3	907	1	VGP3_EBV	P03200 epstein-bar
15	319.5	7.3	860	1	CHI2_COCIM	P54197 coccidioide
16	317	7.2	307	1	SGS3_DROME	P02840 drosophila
17	315	7.2	1161	1	YJ9P_YEAST	P47179 saccharomyc
18	313.5	7.1	886	1	VGP3_EBVA8	Q07284 epstein-bar
19	311	7.1	620	1	EXTN_TOBAC	P13983 nicotiana t
20	307.5	7.0	283	1	EXTN_SORVU	P24152 sorghum vul
21	300.5	6.8	3149	1	TEGU_EBV	P03186 epstein-bar
22	296.5	6.7	634	1	HWP1_CANAL	P46593 candida alb
23	290.5	6.6	1794	1	YAV1_SCHPO	Q10172 schizosacch
24	290	6.6	1162	1	TCNA_TRYCR	P23253 trypanosoma
25	285.5	6.5	1251	1	YQU3_CAEEL	Q09550 caenorhabdi
26	283.5	6.5	2476	1	ZAN_PIG	Q28983 sus scrofa
27	282	6.4	699	1	VGLG_HSV2H	P13290 herpes simp
28	276.5	6.3	338	1	SED1_YEAST	Q01589 saccharomyc
29	270.5	6.2	263	1	SGS3_DROYA	P13728 drosophila
30	269.5	6.1	1083	1	T2D3_HUMAN	O00268 homo sapien
31	268.5	6.1	796	1	YS8A_CAEEL	Q09625 caenorhabdi
32	268	6.1	3421	1	TEGU_HSVEB	P28955 equine herp
33	267.5	6.1	600	1	SP96_DICDI	P14328 dictyosteli
34	266	6.1	721	1	YY21_MYCTU	Q10690 mycobacteri
35	265.5	6.0	901	1	A180_MOUSE	Q61548 mus musculu
36	261.5	5.9	467	1	RPB1_CRIGR	P11414 cricetulus
37	261.5	5.9	1970	1	RPB1_HUMAN	P24928 homo sapien
38	261.5	5.9	1970	1	RPB1_MOUSE	P08775 mus musculu
39	261.5	5.9	439	1	XP2_XENLA	P17437 xenopus lae
40	260	5.9	915	1	A180_RAT	Q05140 rattus norv
41	253.5	5.8	1802	1	HKR1_YEAST	P41809 saccharomyc
42	252.5	5.7	1742	1	GUNA_CALSA	P22534 caldocellum
43	250.5	5.7	817	1	VRP1_YEAST	P37370 saccharomyc
44	249	5.7	1075	1	FLO5_YEAST	P38894 saccharomyc
45	247	5.6	1537	1	FLO1_YEAST	P32768 saccharomyc

ALIGNMENTS

RESULT 1
MUC2_HUMAN
ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
GN MUC2 OR SMUC.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RX MEDLINE; 94132002.
RA GUM J.R. JR., HICKS J.W., TORIBARA N.W., SIDDIKI B., KIM Y.S.;
RL J. BIOL. CHEM. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=COLON;
RX MEDLINE; 93016075.
RA GUM J.R. JR., HICKS J.W., TORIBARA N.W., ROTHE E.-M., LAGACE R.E.,
RA KIM Y.S.;
RL J. BIOL. CHEM. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE; 91358717.
RA TORIBARA N.W., GUM J.R. JR., CULHANE P.J., LAGACE R.E., HICKS J.W.,
RA PETERSEN G.M., KIM Y.S.;
RL J. CLIN. INVEST. 88:1005-1013(1991).
CC -!- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS.
CC -!- SUBUNIT: MULTIMERIC.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC VARIES AMONG DIFFERENT ALLELES.
CC -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC OF SILKWORM HEMOCYTIN.
CC -!- SIMILARITY: CONTAINS 2 VWFC DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR EMBL; L21998; G454154; -.
DR EMBL; M74027; G188864; -.
DR EMBL; M94131; G186396; -.
DR EMBL; M94132; G186398; -.
DR HSSP; P15445; 1PSH.
DR MIM; 158370; -.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01208; VWFC; 2.
KW GLYCOPROTEIN; REPEAT; POLYMORPHISM; SIGNAL.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 5179 MUCIN 2.
FT DOMAIN 1401 1747 APPROXIMATE REPEATS.
FT REPEAT 1401 1416 1.
FT REPEAT 1417 1432 2.
FT REPEAT 1433 1448 3.
FT REPEAT 1449 1464 4.
FT REPEAT 1465 1471 5.
FT REPEAT 1472 1478 6.
FT REPEAT 1479 1494 7A.

```
FT   REPEAT     1495   1517       7B.
FT   REPEAT     1518   1533       8A.
FT   REPEAT     1534   1556       8B.
FT   REPEAT     1557   1572       9A.
FT   REPEAT     1573   1596       9B.
FT   REPEAT     1597   1612       10A.
FT   REPEAT     1613   1635       10B.
FT   REPEAT     1636   1651       11A.
FT   REPEAT     1652   1675       11B.
FT   REPEAT     1676   1683       12.
FT   REPEAT     1684   1699       13.
FT   REPEAT     1700   1715       14.
FT   REPEAT     1716   1731       15.
FT   REPEAT     1732   1747       16.
FT   DOMAIN     4815   4886       VWFC 1.
FT   DOMAIN     4924   4991       VWFC 2.
FT   DOMAIN     5075   5160       CTCK.
FT   DISULFID   5075   5122       BY SIMILARITY.
FT   DISULFID   5089   5136       BY SIMILARITY.
FT   DISULFID   5098   5152       BY SIMILARITY.
FT   DISULFID   5102   5154       BY SIMILARITY.
FT   DISULFID      ?   5159       BY SIMILARITY.
FT   CARBOHYD    163    163       POTENTIAL.
FT   CARBOHYD    423    423       POTENTIAL.
FT   CARBOHYD    670    670       POTENTIAL.
FT   CARBOHYD    770    770       POTENTIAL.
FT   CARBOHYD    894    894       POTENTIAL.
FT   CARBOHYD   1139   1139       POTENTIAL.
FT   CARBOHYD   1154   1154       POTENTIAL.
FT   CARBOHYD   1215   1215       POTENTIAL.
FT   CARBOHYD   1230   1230       POTENTIAL.
FT   CARBOHYD   1246   1246       POTENTIAL.
FT   CARBOHYD   1787   1787       POTENTIAL.
FT   CARBOHYD   1820   1820       POTENTIAL.
FT   CARBOHYD   4339   4339       POTENTIAL.
FT   CARBOHYD   4351   4351       POTENTIAL.
FT   CARBOHYD   4362   4362       POTENTIAL.
FT   CARBOHYD   4373   4373       POTENTIAL.
FT   CARBOHYD   4422   4422       POTENTIAL.
FT   CARBOHYD   4438   4438       POTENTIAL.
FT   CARBOHYD   4502   4502       POTENTIAL.
FT   CARBOHYD   4616   4616       POTENTIAL.
FT   CARBOHYD   4627   4627       POTENTIAL.
FT   CARBOHYD   4752   4752       POTENTIAL.
FT   CARBOHYD   4787   4787       POTENTIAL.
FT   CARBOHYD   4881   4881       POTENTIAL.
FT   CARBOHYD   4888   4888       POTENTIAL.
FT   CARBOHYD   4955   4955       POTENTIAL.
FT   CARBOHYD   4970   4970       POTENTIAL.
FT   CARBOHYD   5019   5019       POTENTIAL.
FT   CARBOHYD   5038   5038       POTENTIAL.
FT   CARBOHYD   5069   5069       POTENTIAL.
FT   CONFLICT   1351   1351       H -> L (IN REF. 3).
FT   CONFLICT   1412   1412       T -> S (IN REF. 3).
FT   CONFLICT   1449   1449       L -> P (IN REF. 3).
FT   CONFLICT   1504   1504       M -> T (IN REF. 3).
FT   CONFLICT   4192   4192       G -> S (IN REF. 2).
SQ   SEQUENCE   5179 AA;  540295 MW;  ECD04582 CRC32;

  Query Match             15.2%;  Score 670;  DB 1;  Length 5179;
  Best Local Similarity   30.6%;  Pred. No. 7.9e-22;
  Matches  220;  Conservative   42;  Mismatches  262;  Indels  196;  Gaps   31;

Qy      91 TLPGPP---------AETTQAPATTQAP----TTTQAPTTTTQAPTTTTQAP---TTTTQ 134
           | | ||         : ||  | || :|    |||  ||||   ||||| :|   ||||
Db    1537 TTPSPPTTTPITPPTSTTTLPPTTTPSPPPTTTTTPPPTTTPSPPTTTTPSPPTITTTTP 1596

Qy     135 APTTTQAP----TTTQAPTTTQAPTTTT--QAPTTTTQAPTTTTQAP----TTTQAPTTT 184
            |||| :|    |||  |||| :| |||    ||:||  | ||| :|    |||  ||||
Db    1597 PPTTTPSPPTTTTTTPPPTTTPSPPTTTPITPPTSTTTLPPTTTPSPPPTTTTTPPPTTT 1656

Qy     185 QAPTTTQAPT-TITQAATTPAATTPAA------------------TTPAA-----TTPA 219
```

```
           :| ||  |:  ||   | |  |||::                    |||::     |||:
Db    1657 PSPPTTTTPSPPITTTTTPPPTTTPSSPITTTPSPPTTTMTTPSPTTTPSSPITTTTTPS 1716

Qy     220 ATT---PAATTPGVPAPTSAP--------VWPP--------ICELLPNGCPADFDIHLLI 260
           :||   |  ||   |:||: |          ||          | |:  | |
Db    1717 STTTPSPPPTTMTTPSPTTTPSPPTTTMTTLPPTTTSSPLTTTPLPPSITPPTFS----- 1772

Qy     261 PHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDE 320
                       |    |      | |  |:    ::     |  |   :   |
Db    1772 ---------------PFSTTTPT----TPCVPLCNWTGWLD-----SGKPNFHKPGGDT 1806

Qy     321 DIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHI 380
           :: |||:   |  ||           |   |     | |
Db    1807 EL-IGDVCGPGWAANISC--------RATMYPDVPIGQL---------------------- 1837

Qy     381 ELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG-CPADFSIDHLLPHESDCGQ 439
                  |::   ||     |  |  |    :  | |  |  | ::: :
Db    1837 ---------GQTVVCDVS----VGLICKNE----DQKPGGVIPMAFCLNYEI-------- 1872

Qy     440 YLQCVHGQTIARPC------------PGNLHFSPATQSCESPVTAGCQVFECDSDNQCTS 487
           :||   : : :|             |     :  |    :|   | |     :    |:
Db    1872 NVQCC--ECVTQPTTMTTTTTENPTPPTTTPITTTTTVTPTPTPTGTQT---PTTTPITT 1926

Qy     488 TAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVP--PTTAIPTPAPTAAPTA 544
           |   | ||   |  ||  |    :|| |  ||  |  |       ||   || ||   |
Db    1927 TTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQ 1986

Qy     545 APTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAV 604
           |||    : ||||  ||   |  |||       |||  ||   |  ||  |    ||
Db    1987 TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVT 2046

Qy     605 PEIPTTVTSPPTAAP----TTAAPAPNTTVTVPPTAAP----TTAAPAPNTTVTAPPTAA 656
           |    | |  ||  |    ||  | |  | |  ||  |    ||  | |  | |  ||
Db    2047 PTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTT 2106

Qy     657 P----TTAAPAPNTTVTVPPTAAP-----TAAP---PTVAHAPNTTAAPVTTTSAPATTP 704
           |    ||  | |  | |  ||  |     |  |   ||    | ||    ||| |  ||
Db    2107 PITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTP 2166
```

```
RESULT   2
CPN_DROME
ID   CPN_DROME      STANDARD;      PRT;   865 AA.
AC   Q02910;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE   CALPHOTIN.
GN   CPN OR CAP.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE; 93165729.
RA   MARTIN J.H., BENZER S., RUDNICKA M., MILLER C.A.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 90:1531-1535(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE; 93165730.
RA   BALLINGER D.G., XUE N., HARSHMAN K.D.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 90:1536-1540(1993).
CC   -!- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
CC       REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
CC       OF CA+2 PER MOL OF PROTEIN.
CC   -!- SUBUNIT: HOMODIMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC   -!- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
CC       COMPOUND EYES AND OCELLI.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
CC       DEVELOPMENT.
DR   EMBL; L02111; G157032; -.
```

```
DR   EMBL; L05080; G157072; -.
DR   PIR; A47282; A47282.
DR   FLYBASE; FBgn0010218; Cpn.
KW   CALCIUM-BINDING.
FT   CONFLICT     36     36       A -> AVAPAVVA (IN REF. 2).
FT   CONFLICT     43     43       I -> T (IN REF. 2).
FT   CONFLICT     64     64       I -> V (IN REF. 2).
FT   CONFLICT     76     76       T -> A (IN REF. 2).
FT   CONFLICT    100    100       P -> PP (IN REF. 2).
FT   CONFLICT    126    127       VQ -> AP (IN REF. 2).
FT   CONFLICT    154    154       I -> V (IN REF. 2).
FT   CONFLICT    160    160       S -> T (IN REF. 2).
FT   CONFLICT    534    534       A -> E (IN REF. 2).
FT   CONFLICT    699    699       I -> T (IN REF. 2).
FT   CONFLICT    703    703       V -> L (IN REF. 2).
FT   CONFLICT    721    721       D -> E (IN REF. 2).
SQ   SEQUENCE   865 AA;  84781 MW;  7C1F27FB CRC32;

  Query Match             10.4%;  Score 457;  DB 1;  Length 865;
  Best Local Similarity   24.2%;  Pred. No. 2.2e-13;
  Matches  197;  Conservative   46;  Mismatches  250;  Indels  320;  Gaps   35;

Qy      91 TLPGPPAETTQAPAT---TQAPTTTQAPTTTTQAPT-------------TTTQAPTTTT 133
           |:| | :    || |     ||    :|     ||                : |  | |
Db       5 TIPSPVSAPVAAPVTPSAVAAPVQVVSPAAVAPAPAAPIAVTPVAPPPTLASVQPATVTI 64

Qy     134 QAPTTTQA----PTTTQAPTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTQAPTT 189
            ||    |   | : ||    |||    :| :|   |    | |    ||    ||
Db      65 PAPAPIAAASVTPVASVAPPVVAAPTPPAASPVST---PVAVAQIPVAVSAPV---APPV 118

Qy     190 TQAPTTITQ-AATTPAATTP----AATTPAATTPAATTPAATTPGVPAPTSAPVWPPICE 244
              || : |     |   ||    :| |||| ||  :   |: | ||| |: ||  |:
Db     119 AATPTPVVQIPVAAPVIATPPVAASAPTPAAVTPVISPVIASPPVVPANTTVPVAAPVA- 178

Qy     245 LLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECD 304
            :|   |                                    | :    :||
Db     178 AVPAAVPV-----------------------------------VAPVLAPAVAPAVAPVV 202

Qy     305 GEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRC 364
            |    ||| |     | :  : :   |       |:|  :
Db     203 AETPAPPPVAE------IPVATIPECVAP-------LIPEVS------------------ 232

Qy     365 GAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPAD 424
                          |  | |   :| | |   |       ||       |   ||
Db     232 ------------------VVATKPLAAAEPVVVAPPA-------TE------TPVVAPAA 260

Qy     425 FSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQ 484
            |     || |                |       :| : | | || |
Db     261 AS-----PHVS--------------VAPAVETAVVAPVSASTEPPVAA------------ 290

Qy     485 CTSTAAPTAAPTAAPTAA-------------PTAAPTAAPSTVVPP-------------- 518
            | | ||   |  ||  |             || ||   | |: ||
Db     290 ATLTTAP-ETPALAPVVAESQVAANTVVATPPTPAP--EPETIAPPVVAETPEVASVAVA 346

Qy     518 -ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVT-----VPPTAAPTAAPTT 571
             |||     |||  |   |||  | |  | | |   |          :||  ||:  | :
Db     347 ETTPPV----VPPVAAESIPAPVVATTPVPATLAVTDPDVTASAVPELPPVIAPSPVP-S 401

Qy     572 AVPEIPITV----------------------TSAPTAAPT-------------AAPTA 594
           || | |: :                        | || :||             |||:
Db     402 AVAETPVDLAPPVLPPVAAEPVPAVVAEETPETPAPASAPVTIAALDIPEVAPVIAAPSD 461

Qy     595 APTAAPTTAVPEI---PTTVTSPPTAAPTTAAP-----------------APNTTVT-- 632
           ||  ||: | | :   ||| : | | ||  | |                 ||   ||
Db     462 APAEAPSAAAPIVSTPPTTASVPETTAPPAAVPTEPIDVSVLSEAAIETPVAPPVEVTTE 521

Qy     632 ------VPPTAA-------------------------------------PTTAAPAPNTT 648
                  || ||                                     | |: | ||
Db     522 VAVADVAPPEAAADLIIEPVEPPAPIPDLLEQTTSVPAVEAAESTSSPIPETSLPPPNEA 581

Qy     649 VTAPPTA-APTTA---APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPV--TTTSAPAT 702
           | :|  | || ||    | |  ::  |    |  ||  :  | :    ||  |:|| | |
```

```
Db     582 VASPEVAVAPITAPEPIPEPEPSLATPTEPIPVEAPVVIQEAVDAVEVPVTETSTSIPET 641

Qy     703 TPE------DDDIDPPLPNDPI---NPCVEECN 726
           | |      :  :|| :   |:    | |   |
Db     642 TVEFPEAVAEKVLDPAITEAPVTTQEPDVANIN 674


RESULT   3
VGLX_HSVEB
ID   VGLX_HSVEB     STANDARD;      PRT;   797 AA.
AC   P28968;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE   GLYCOPROTEIN X PRECURSOR.
GN   71.
OS   EQUINE HERPESVIRUS TYPE 1 (STRAIN AB4P) (EHV-1).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92295566.
RA   TELFORD E.A.R., WATSON M.S., MCBRIDE K., DAVISON A.J.;
RL   VIROLOGY 189:304-316(1992).
DR   EMBL; M86664; G330862; -.
DR   PIR; H36802; VGBEX1.
KW   GLYCOPROTEIN; TRANSMEMBRANE; SIGNAL.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    797       GLYCOPROTEIN X.
FT   DOMAIN       23    465       SER/THR-RICH.
FT   TRANSMEM    766    790       POTENTIAL.
FT   CARBOHYD    590    590       POTENTIAL.
SQ   SEQUENCE   797 AA;  80342 MW;  18DE40E1 CRC32;

  Query Match             10.2%;  Score 446.5;  DB 1;  Length 797;
  Best Local Similarity   25.3%;  Pred. No. 5.7e-13;
  Matches  182;  Conservative   41;  Mismatches  186;  Indels  311;  Gaps   20;

Qy      78 SAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTQAPTT--TTQA 135
           | |:   || :  : |   : ||  | :|   |||  || :|  ||||| |||  || |
Db      77 STQSSSTAATSS-SAPSTASSTTSIPTSTSTETTTTTPTASTTTPTTTTAAPTTAATTTA 135

Qy     136 PTT---TQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTT--------------------TT 172
            ||    | | |||   | |  |||||   |||| | ||                    ||
Db     136 VTTAASTSAETTTATATATSTPTTTTPTSTTTTTATTTVPTTASTTTDTTTAATTTAATT 195

Qy     173 QAPTTTQAPT---TTQAPTTTQAPTT------ITQAATTPAATTPAATTPAATTPAATTP 223
           | |||  | |   || | ||| | ||       | |||| |||| |||| |||| ||||
Db     196 TAATTTAATTTAATTTAATTTAATTTAATTSSATTAATTTAATTTAATTTAATTTAATTT 255

Qy     224 AATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCP 283
           |||| |  :|||
Db     256 AATTTG--SPTS---------------------------------------------- 266

Qy     284 EGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLP 343

Db     266 ------------------------------------------------------------ 266

Qy     344 HGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACT 403
                                                  | |   | |
Db     266 ----------------------------------------GSTSTTGASTSTP------- 279

Qy     404 GWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPAT 463
                                                                  | :|
Db     279 ----------------------------------------------------SAST 282

Qy     464 QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPA- 523
           :  :|               |||:   || |: ||  ||:| |:| ||   | : |
Db     283 ATSATP---------------TSTSTSAAATTSTPT--PTSAATSAESTTEAPTSTPTT 324

Qy     523 -TAAPVPPTTAIPTPAPT---AAPTAAPTTA-------APESPT--TVTVPPTAAPTAAP 569
            |  |   |||  :|  |   |: |:| |||       :|:|  |  | |  |::  |  |
Db     325 DTTTPSEATTATTSPESTTVSASTTSATTTAFTTESHTSPDSSTGSTSTAEPSSTFTLTP 384
```

```
Qy    570 TTAVPEI-PITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNT 628
          :|| |     | :|| | : :   :  ||    :     ||  |    : |  :
Db    385 STATPSTDQFTGSSASTESDSTDSSTVPTTGTESITESSSTTEASTNLGSSTYESTEALE 444

Qy    629 TVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTA---------- 675
          |     |:  || :|:| |   |     |       :||:    ||
Db    445 TPDGNTTSGNTTPSPSPRTPSFADTQQTPDNGVSTQHTTINDHTTANAQKHAGHHRGRAG 504

Qy    675 APTAAPPTVAHAPNTTAAPVTTTSAPATTPEDD----------DIDPPLPNDPINPCVE 723
          :|   :|   ||  |   | :|  | :||          :| | ||  |  | :|
Db    505 GRRGSPQGGSH---TTPHPDRLTPSPDDTYDDDTNHPNGRNNSIEIVPQLP--PDRPIIE 559
```

```
RESULT   4
MUC1_XENLA
ID   MUC1_XENLA     STANDARD;      PRT;   662 AA.
AC   Q05049;
DT   01-OCT-1994 (REL. 30, CREATED)
DT   01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   INTEGUMENTARY MUCIN C.1 (FIM-C.1) (FRAGMENT).
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKIN;
RX   MEDLINE; 93077556.
RA   HAUSER F., HOFFMANN W.;
RL   J. BIOL. CHEM. 267:24620-24624(1992).
CC   -!- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC       INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: SKIN.
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC   -!- ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN
CC       MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS 6 P-TYPE (TREFOIL) DOMAINS.
DR   EMBL; L02115; G951460; -.
DR   PIR; A45155; A45155.
DR   HSSP; P04002; 1ATF.
DR   PROSITE; PS00025; P_TREFOIL; 6.
KW   REPEAT; AMPHIBIAN SKIN; GLYCOPROTEIN; ALTERNATIVE SPLICING.
FT   NON_TER       1      1
FT   DOMAIN       81    144       8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT                                ALA/THR-RICH.
FT   REPEAT       81     88       1-1.
FT   REPEAT       89     96       1-2.
FT   REPEAT       97    104       1-3.
FT   REPEAT      105    112       1-4.
FT   REPEAT      113    120       1-5.
FT   REPEAT      121    128       1-6.
FT   REPEAT      129    136       1-7.
FT   REPEAT      137    144       1-8.
FT   DOMAIN      161    202       P-TYPE 1.
FT   DOMAIN      218    301       8 X APPROXIMATE TANDEM REPEATS, THR-RICH.
FT   REPEAT      218    224       2-1.
FT   REPEAT      225    239       2-2.
FT   REPEAT      240    249       2-3.
FT   REPEAT      250    259       2-4.
FT   REPEAT      260    275       2-5.
FT   REPEAT      276    287       2-6.
FT   REPEAT      288    294       2-7.
FT   REPEAT      295    301       2-8.
FT   DOMAIN      306    347       P-TYPE 2.
FT   DOMAIN      353    394       P-TYPE 3.
FT   DOMAIN      402    522       12 X APPROXIMATE TANDEM REPEATS,
FT                                THR-RICH.
FT   REPEAT      402    411       3-1.
FT   REPEAT      412    419       3-2.
FT   REPEAT      420    431       3-3.
FT   REPEAT      432    443       3-4.
FT   REPEAT      444    453       3-5.
FT   REPEAT      454    460       3-6.
FT   REPEAT      461    472       3-7.
FT   REPEAT      473    479       3-8.
FT   REPEAT      480    491       3-9.
FT   REPEAT      492    498       3-10.
FT   REPEAT      499    515       3-11.
FT   REPEAT      516    522       3-12.
FT   DOMAIN      525    566       P-TYPE 4.
FT   DOMAIN      572    613       P-TYPE 5.
FT   DOMAIN      620    661       P-TYPE 6.
FT   DISULFID    162    188       BY SIMILARITY.
FT   DISULFID    172    187       BY SIMILARITY.
FT   DISULFID    182    199       BY SIMILARITY.
FT   DISULFID    307    333       BY SIMILARITY.
FT   DISULFID    317    332       BY SIMILARITY.
FT   DISULFID    327    344       BY SIMILARITY.
FT   DISULFID    354    380       BY SIMILARITY.
FT   DISULFID    364    379       BY SIMILARITY.
FT   DISULFID    374    391       BY SIMILARITY.
FT   DISULFID    526    552       BY SIMILARITY.
FT   DISULFID    536    551       BY SIMILARITY.
FT   DISULFID    546    563       BY SIMILARITY.
FT   DISULFID    573    599       BY SIMILARITY.
FT   DISULFID    583    598       BY SIMILARITY.
FT   DISULFID    593    610       BY SIMILARITY.
FT   DISULFID    621    647       BY SIMILARITY.
FT   DISULFID    631    646       BY SIMILARITY.
FT   DISULFID    641    658       BY SIMILARITY.
FT   VARIANT     276    276       K -> E.
FT   VARIANT     354    354       C -> R.
FT   VARIANT     415    415       T -> A.
SQ   SEQUENCE   662 AA;  67774 MW;  00C98798 CRC32;
```

Query Match 9.8%; Score 430.5; DB 1; Length 662;
Best Local Similarity 22.4%; Pred. No. 2.3e-12;
Matches 194; Conservative 27; Mismatches 226; Indels 419; Gaps 28;

```
Qy     65 APRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTT----QAPTT--- 118
          ||   | |     :| |    | ||     |   | :|||  ||  |    :||||
Db     73 APATTAAGKAPTTAAATAPTTAAAG----APTTATGKAPATAAAPVPTTAASKAPTTAAA 128

Qy    118 ------TTQAPTTTTQA---------------------------------------- 129
                 ||||   |
Db    129 ATHSTAAAAAPTTAASAAKSKERSTSSSSEEEHCHVKPSKREMCGSKGITKKQCKKKNCC 188

Qy    129 ----------------------PTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTT 164
                               ||||:|||| |  |||  ||||   |  |   |||
Db    189 FDPKGHGGIHCFHRKPKGHSHEEHTTTTTKAPTTIQIATTTTTPTTTTTTTKATPTTTTT 248

Qy    165 TQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPA 224
          |:|  |||   |||:| |||  ||||   || |:| |||  ||   |||  ||  |||
Db    249 TKATPTTT---TTTKATTTTTTPTTT---TTTTKATTTPTTTT--TTTPTTTTTKATTTT 300

Qy    225 ATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPE 284
           || |
Db    301 TTTSG---------------------------------------------------- 306

Qy    285 GLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPH 344
                      ||  | |
Db    306 ----------------ECKMEPS------------------------------------ 313

Qy    345 GNRCDKYYQCVHGNLVERRCGA-GTHFSFELQQ---CDHIELVGCTLPGGESEEVDVDED 400
          |   | :  : | :|   |  |  :  |   | :      ||    |:  |
Db    313 -----KREDCGYSGITESQCRTKGCCFDSSIPQTKWCFY------TL----SQVAD---- 354

Qy    401 ACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFS 460
            |  || :                    |||
Db    354 ------CKVEPSQ-------------------RVDCG---------------------- 366

Qy    461 PATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATP 520
                  |    :|   | |  ::      :       : :  ||  |   | |
Db    366 ----------FRGITADQCRQKNCCFDSSI-----SGTKWCFYSTSQVAATKTTTTPTT- 410
```

```
Qy    521 PATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITV 580
                  |||   | | |  ||   ||     ||| |
Db    410 -----TTTPTTTTTTKATTTPTTTTTT-----PTTTTT-------------------- 439

Qy    581 TSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTT 640
                  |   | | |  |||  |   || |   |  |||    | || |   |  |||
Db    439 --------TTTTTKATTTPTTTTP--TTTTTKATTTTPTTTTTTPTTTTTKATTTTPTT 488

Qy    641 AAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAP 700
            | || |   |  |||      || |                    || |  ||||
Db    489 TTTTPTTTTTKATTTTPTT------TTTTT----------------TTTKATTTTTSGE 525

Qy    701 ATTPEDDDIDPPLPNDPINPC----------------------VEECNVLPWAHADC-- 736
                  |   |    : |                     | :| | | :  ||
Db    526 CKMEPSKRADCGYPGITESQCRSKGCCFDSSIPQTKWCFYSLPQVADCKVAPSSRVDCGF 585

Qy    736 --------------------DKYWVCDGNNQVLVVCSEGLQFNPTTKTCD---------F 766
                               |     :|   :||      | ||  |
Db    586 GGITADQCRQRNCCFDSSISGTKWCFYSTSQGNAMCS-----GPPTKRRDCGYPGISSSV 640

Qy    767 ACNVGCVRSNIQMSESYEGVQVFIPW 792
            | ||   |  |:         :||
Db    641 CINRGCCWDNSVMN---------VPW 657
```

```
RESULT   5
MUC1_HUMAN
ID   MUC1_HUMAN     STANDARD;      PRT;  1255 AA.
AC   P15941; P15942; P13931; P17626;
DT   01-JAN-1990 (REL. 13, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)
DE   (EPISIALIN) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
DE   (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG)
DE   (PEANUT-REACTIVE URINARY MUCIN) (PUM).
GN   MUC1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PANCREAS;
RX   MEDLINE; 90368716.
RA   LAN M.S., BATRA S.K., QI W.-N., METZGAR R.S., HOLLINGSWORTH M.A.;
RL   J. BIOL. CHEM. 265:15294-15299(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90202794.
RA   LIGTENBERG M.J.L., VOS H.L., GENNISSEN A.M.C., HILKENS J.;
RL   J. BIOL. CHEM. 265:5573-5578(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST CARCINOMA;
RX   MEDLINE; 90368715.
RA   GENDLER S.J., LANCASTER C.A., TAYLOR-PAPADIMITRIOU J., DUHIG T.,
RA   PEAT N., BURCHELL J., PEMBERTON L., LALANI E.-N., WILSON D.;
RL   J. BIOL. CHEM. 265:15286-15293(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91097524.
RA   LANCASTER C.A., PEAT N., DUHIG T., WILSON D.,
RA   TAYLOR-PAPADIMITRIOU J., GENDLER S.J.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 173:1019-1029(1990).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST CARCINOMA;
RX   MEDLINE; 90276413.
RA   WRESCHENER D.H., HAREUVENI M., TSARFATY I., SMORODINSKY N.,
RA   HOROV J., ZARETSKY J., KOTKES P., WEISS M., LATHE R., DION A.,
RA   KEYDAR I.;
RL   EUR. J. BIOCHEM. 189:463-473(1990).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST CARCINOMA;
RX   MEDLINE; 90276414.
RA   HAREUVENI M., TSARFATY I., ZARETSKY J., KOTKES P., HOREV J.,
RA   ZRIHAN S., WEISS M., GREEN S., LATHE R., KEYDAR I., WRESCHNER D.H.;
RL   EUR. J. BIOCHEM. 189:475-486(1990).
RN   [7]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91033045.
RA   TSARFATY I., HAREUVENI M., HOREV J., ZARETSKY J., WEISS M.,
RA   JELTSCH J.M., GARNIER J.M., LATHE R., KEYDAR I., WRESCHNER D.H.;
RL   GENE 93:313-318(1990).
RN   [8]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE; 88330762.
RA   GENDLER S.J., TAYLOR-PAPADIMITRIOU J., DUHIG T., ROTHBARD J.,
RA   BURCHELL J.;
RL   J. BIOL. CHEM. 263:12820-12823(1988).
RN   [9]
RP   SEQUENCE OF 1-46 FROM N.A.
RC   TISSUE=BREAST CARCINOMA;
RA   BULUWELA L., LIU Q., LUQMANI Y.A., GOMM J.J., COOMBES R.C.;
RL   SUBMITTED (OCT-1992) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC       CYTOSKELETON.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
CC       IS ALSO PRODUCED.
CC   -!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL
CC       TUMORS, SUCH AS BREAST CANCER.
CC   -!- PTM: HIGHLY GLYCOSYLATED (N- & O-LINKED CARBOHYDRATES & SIALIC
CC       ACID).
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC       VARIES AMONG DIFFERENT ALLELES. THE BREAST TUMOR MUC1 HAS
CC       GENERALLY ONLY 5 TO 7 REPEATS.
CC   -!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE
CC       SPLICING OF THE MUC1 GENE.
DR   EMBL; J05582; G189599; -.
DR   EMBL; M32738; G182124; -.
DR   EMBL; M32739; G182129; -.
DR   EMBL; J05581; G188870; -.
DR   EMBL; M61170; G189784; -.
DR   EMBL; X52229; G37054; ALT_SEQ.
DR   EMBL; X52228; G36435; ALT_SEQ.
DR   EMBL; M35093; G182253; ALT_SEQ.
DR   EMBL; Z17324; G34807; -.
DR   EMBL; Z17325; G34809; -.
DR   EMBL; M21868; G386963; ALT_SEQ.
DR   PIR; A35175; A35175.
DR   PIR; B35175; B35175.
DR   PIR; S10218; S10218.
DR   MIM; 158340; -.
KW   GLYCOPROTEIN; SIGNAL; CYTOSKELETON; ACTIN-BINDING; TRANSMEMBRANE;
KW   REPEAT; ALTERNATIVE SPLICING; POLYMORPHISM.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24   1255       MUCIN 1.
FT   DOMAIN       24   1162       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1163   1186       POTENTIAL.
FT   DOMAIN     1187   1255       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       81    960       44 X 20 AA TANDEM REPEATS.
FT   CARBOHYD    957    957       POTENTIAL.
FT   CARBOHYD    975    975       POTENTIAL.
FT   CARBOHYD   1029   1029       POTENTIAL.
FT   CARBOHYD   1055   1055       POTENTIAL.
FT   CARBOHYD   1133   1133       POTENTIAL.
FT   VARSPLIC     19     19       T -> ATTAPKPAT (IN FORM B).
FT   VARSPLIC     20     22       MISSING (IN FORM C).
FT   VARSPLIC     20     31       MISSING (IN FORM D).
FT   VARSPLIC    126    905       MISSING (IN POLYMORPHIC EPITHELIAL
FT                                VARIANT).
FT   VARSPLIC   1077   1087       FLQIYKQGGFL -> VSIGLSFPMLP (IN SECRETED
FT                                FORM).
FT   VARSPLIC   1088   1255       MISSING (IN SECRETED FORM).
```

```
FT   CONFLICT    1021   1021       S -> T (IN REF. 3).
FT   CONFLICT    1251   1251       A -> T (IN REF. 3).
SQ   SEQUENCE    1255 AA;  122072 MW;  B06B1004 CRC32;


  Query Match             9.7%;  Score 428;  DB 1;  Length 1255;
  Best Local Similarity   26.7%;  Pred. No. 4.8e-12;
  Matches  208;  Conservative   45;  Mismatches  294;  Indels  232;  Gaps   37;

Qy      17 RPEVSDAEKNPALHE--PHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRDCAPGTE 74
           || :      |  |:    ||  ||      |          :|:        |||:
Db     105 RPALGST--TPPAHDVTSAPDNKPAPGSTAPPA----------HGVTSAPDTRPAPGS- 151

Qy      75 FKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTT-------TQAPTTTTQ 127
             :|       |  | | | :    |   | || |  || :|       | || |
Db     151 ---TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 207

Qy     128 APTTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTTQAPTTTTQAPTTTQAPTTTQA 186
           :|   |   | || |  || :|  |     | || |     :|   |   | || |  |
Db     208 PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA 267

Qy     187 PTTTQAPT-TITQAATT---PAATTPAA---TTPAATTPAATTPAATTPGVPAPTSAP-- 238
           | :|  |   :| |  |   | :| | |   |:   | ||   | :| |     ||||
Db     268 PGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPA---PGSTAPPAHGVTSAPDT 324

Qy     238 ------VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTF--EQRCPEGLYFN 289
                 ||                                ::| |   : |   |
Db     325 RPAPGSTAPP------------------------------AHGVTSAPDTRPAPGSTAP 353

Qy     290 PYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCD 349
           |      :|      |  | |||
Db     354 PAHGVTSAPDTRPAPG--STAPP----------------------------------- 375

Qy     350 KYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP- 409
             ||                |     | | |          |    |   |
Db     375 -----AHG--------------VTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 414

Qy     409 ----TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA---RPCPGNLHFSP 461
               |   :  | |                   |      || | |   || ||:    |
Db     415 AHGVTSAPDTRPAP------------------GSTAPPAHGVTSAPDTRPAPGST-APP 454

Qy     462 ATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVV--PPAT 519
           |     :|          |:     ||| |    |:||    |    || |:  |   | |
Db     455 AHGVTSAP----------DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 504

Qy     520 PPATAAPVPP---TTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEI 576
             ||  :  ||    |: |   |    || |      :| |   | : || |   |: |:
Db     505 RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 564

Qy     577 -PITVTSAP-----TAAPTAAPTAAPTAAPTTAVPEIPTT-----VTSPP----TAAPTT 621
            |   ::||     |:||   |    || |   |   | |       |:||    |:|| |
Db     565 RPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDT 624

Qy     622 AAPAPNTTVTVPP----TAAPTTAAPAPNTTVTAPP----TAAPTTAAPAPNTTVTVPPT 673
             ||| :  | ||    |:|| |  ||| :  ||||    |:|| |  ||| :  | ||
Db     625 -RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPAHGVTSAPDT-RPAPGS--TAPPA 675

Qy     674 AAPTAAPPT-------------VAHAPNTTAAPVTT-------TSAPATTPEDDDIDPP 712
             |:|| |             |  ||:|  || :|        |||| | |      ||
Db     676 HGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPP 734


RESULT   6
AMYH_YEAST
ID   AMYH_YEAST     STANDARD;      PRT;  1367 AA.
AC   P08640; P08068;
DT   01-AUG-1988 (REL. 08, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   GLUCOAMYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-
DE   GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).
GN   STA1 OR STA2 OR MAL5 OR YIR019C.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA   CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
RA   GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA   LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA   RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA   WALSH S.V., WHITEHEAD S.;
RL   SUBMITTED (DEC-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX   MEDLINE; 87194600.
RA   YAMASHITA I., NAKAMURA M., FUKUI S.;
RL   J. BACTERIOL. 169:2142-2149(1987).
RN   [3]
RP   SEQUENCE OF 1-31 FROM N.A.
RC   STRAIN=SPX101-1C;
RX   MEDLINE; 89031230.
RA   PARDO J.M., IANEZ E., ZALACAIN M., CLAROS M.G., JIMENEZ A.;
RL   FEBS LETT. 239:179-184(1988).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
CC       GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
CC       WITH RELEASE OF BETA-D-GLUCOSE.
DR   EMBL; Z47047; G763364; -.
DR   EMBL; Z38061; G557822; -.
DR   EMBL; M16164; G172525; -.
DR   EMBL; M16165; G172526; -.
DR   EMBL; X13857; G4552; -.
DR   PIR; B26877; B26877.
DR   PIR; A26877; A26877.
DR   PIR; S48478; S48478.
DR   SGD; L0002104; STA1.
KW   HYDROLASE; GLYCOSIDASE; POLYSACCHARIDE DEGRADATION; GLYCOPROTEIN;
KW   SIGNAL; MULTIGENE FAMILY.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1367       GLUCOAMYLASE S1/S2.
FT   DOMAIN      210   1367       SER/THR-RICH.
FT   CARBOHYD    817    817       POTENTIAL.
FT   CARBOHYD    874    874       POTENTIAL.
SQ   SEQUENCE   1367 AA;  136110 MW;  212ED195 CRC32;


  Query Match             9.3%;  Score 410;  DB 1;  Length 1367;
  Best Local Similarity   19.9%;  Pred. No. 3e-11;
  Matches  152;  Conservative  124;  Mismatches  238;  Indels  248;  Gaps   28;

Qy      32 PHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCT 91
           || |  |            |||        |    : |   |       :     :
Db     288 PHHDTTP-----------CTK-------KKTTTSKTCTKKTTTPVPTPSSSTTESSSAP 328

Qy      92 LPGPPAETTQ---APATTQAPTTTQAPTTTQAPTT------TTQAPTTTTQAPTTTQAP 142
           :| | : ||:   || |:    :: ||  |  : ||       | : | :: || |:
Db     329 VPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTT 388

Qy     143 TTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTT------TQAPTTTQAPTTTQAPTT- 196
             :: ||  | : :||  :    | : | :: || |      : || |:    :: || |
Db     389 ESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTS 448

Qy     196 -ITQAATTPAATTPAATTPAATTP--AATTPAATTPGVPAP-------TSAPVWPPICEL 245
            |:::: |  |  ::|| ::: |  ::|| ::: | || |        :||||
Db     449 STTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAP-VPTPSSSTTESSSAPV------- 501

Qy     246 LPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDG 305
                                   ::  |     |             :|::   :
Db     501 --------------------------TSSTTESSSAP----------VPTPSSSTTES 522

Qy     306 EISPAP-PVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRC 364
             :||| | :   |     :
Db     523 SSAPAPTPSSSTTESSSAPV---------------------------------------- 543

Qy     365 GAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPAD 424
```

```
Db    543 ---------------------TSSTTESSSAPV-----------------PTPSSSTTE 563

Qy    425 FSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQ 484
Db    564 SS----------------------STPVT-----SSTTESSSAPVPT-------PSSST 588

Qy    485 CTSTAAPTAAP-------TAAPTAAPTAAPTAAPSTVVPPATPPATAAPVP-------PT 530
Db    589 TESSSAPVPTPSSSTTESSSAPAPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTES 648

Qy    531 TAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPIT-VTSAPTAAPT 589
Db    649 SSAPVPTPSSSTTESSSAPVP-TPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPV 707

Qy    590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVT----VPPTAAPTTAAPAP 645
Db    708 PTPSSSTTESSSAPVP-TPSSSTTESSSAP---VPTPSSSTTESSSAPVTSSTTESSSAP 763

Qy    646 NTTVTAPPTAAPTTAAPAPNTTVT----VP-PTAAPTAAPPTVAHAPNTTAAPVTTTSAP 700
Db    764 VPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSVAPVPTPSSSSNITSSAP 823

Qy    701 ATTP-----EDDDIDPPLPND--------PINPCVEECNVLP 729
Db    824 SSTPFSSSTESSSVPVPTPSSSTTESSSAPVSSSTTESSVAP 865
```

```
RESULT   7
SLP1_CLOTM
ID   SLP1_CLOTM     STANDARD;      PRT;  1664 AA.
AC   Q06852;
DT   01-JUN-1994 (REL. 29, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER
DE   PROTEIN 1).
GN   OLPB.
OS   CLOSTRIDIUM THERMOCELLUM.
OC   PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCIB 10682;
RX   MEDLINE; 93209931.
RA   FUJINO T., BEGUIN P., AUBERT J.-P.;
RL   J. BACTERIOL. 175:1891-1899(1993).
CC   -!- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC   -!- SUBCELLULAR LOCATION: CELL WALL.
CC   -!- SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.
DR   EMBL; X67506; G296881; -.
DR   PROSITE; PS01072; SLH_DOMAIN; 2.
KW   CELL WALL; S-LAYER; SIGNAL; REPEAT.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29   1664       CELL SURFACE GLYCOPROTEIN 1.
FT   DOMAIN       36    763       4 X 156 AA APPROXIMATE REPEATS.
FT   REPEAT       36    191       1.
FT   REPEAT      207    363       2.
FT   REPEAT      409    565       3.
FT   REPEAT      607    763       4.
FT   DOMAIN      771   1377       APPROXIMATE TANDEM REPEATS OF
FT                                T-P-S-D-E-P.
FT   DOMAIN     1378   1449       GLY/PRO/SER/THR-RICH.
FT   DOMAIN     1453   1494       SLH 1 (INCOMPLETE).
FT   DOMAIN     1495   1565       SLH 2.
FT   DOMAIN     1566   1625       SLH 3.
FT   DOMAIN     1626   1646       SLH 4 (INCOMPLETE).
SQ   SEQUENCE   1664 AA;  178194 MW;  BFEC7D6A CRC32;
```

```
  Query Match             8.5%;  Score 372.5;  DB 1;  Length 1664;
  Best Local Similarity   23.5%;  Pred. No. 1.3e-09;
  Matches  176;  Conservative   66;  Mismatches  325;  Indels  183;  Gaps   30;
```

```
Qy     18 PEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRD-CAPGTEFK 76
Db    825 PTPSD-EPTPS-DEPTPSDEPTPSDEPTPSETP-------EEPIPTDTPSDEPTPSDE-- 874

Qy     77 FSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAP 136
Db    874 ----------------PTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEP 917

Qy    137 TTTQAPTTTQAPTTTQAPTTT----TQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQA 192
Db    918 TPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 977

Qy    193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPG-VPAPTSAPVWPPICELLPNGCP 251
Db    978 PTPSDE--PTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPT--PSDEPTPSDEP 1033

Qy    252 ADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAP 311
Db   1034 TPSDEP--TPSD-------EPTPSETPEEPIP-----------TDTPSD-----EPTPSD 1068

Qy    312 PVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFS 371
Db   1069 EPTPSDEPTPSDEPTPSDEPTPSETP-EEPIPTDTPSD---------------------- 1106

Qy    372 FELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLL 431
Db   1106 ---------EPTPSDEPTPSDEPTPSDEPTPSDEPTPSETPE-EPIPTDTPSD------E 1149

Qy    432 PHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAP 491
Db   1150 PTPSD--------------EPTPSD---EPTPSDEPTP----------SDEPTPSETPE 1181

Qy    492 TAAPTAAPTAAPTAAPTAAPSTVVPPATPPATA-APVP---PTTAIPTPAPTAAPTAAPT 547
Db   1182 EPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDE 1241

Qy    548 TAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPE 606
Db   1242 PTPSDEPTPSDEPTPSDEPTPSET---PEEPIP-TDTPSDEPTPSDEPTPSDEPTPSDEP 1297

Qy    607 IPTTVTSP---------PTAAPTTAAPAPNTTVTVPPTAA--------------PTTA-- 642
Db   1298 TPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1357

Qy    642 -APAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAP--------------PTVAHA 686
Db   1358 PTPSDEPTPSDEPTPSETPEEPTPTTTPTPTPSTTPTSGSGGSGGSGGGGGGGGGTVPTS 1417

Qy    687 PNTTAAPVTTTSAPATTPEDDDIDPPLPND 716
Db   1418 PTP-----TPTSKPTSTPAPTEIEEPTPSD 1442
```

```
RESULT   8
ANP_NOTCO
ID   ANP_NOTCO      STANDARD;      PRT;   822 AA.
AC   P24856;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   ANTIFREEZE GLYCOPEPTIDE POLYPROTEIN AFGP7/AFGP8 PRECURSOR.
GN   AFGP8.
OS   NOTOTHENIA CORIICEPS NEGLECTA (BLACK ROCKCOD) (YELLOWBELLY ROCKCOD).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;
OC   OSTEICHTHYES; ACTINOPTERYGII; PERCIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91067687.
RA   HSIAO K.-C., CHENG C.-H.C., FERNANDES I.E., DETRICH H.W. III,
RA   DEVRIES A.L.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 87:9265-9269(1990).
CC   -!- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED BY THE LIVER AND SECRETED INTO
```

```
CC       THE BLOOD FROM WHICH THEY BECOME DISTRIBUTED TO ALMOST THE ENTIRE
CC       EXTRACELLULAR SPACE.
CC   -!- PTM: THE DISACCHARIDE GALACTOSE-N-ACETYLGALACTOSAMINE IS ATTACHED
CC       TO THREONINES IN AFGP8 AND AFGP7.
CC   -!- CONTAINS 44 COPIES OF AFGP8 AND TWO COPIES OF AFGP7.
DR   EMBL; M55000; G213392; -.
DR   PIR; A38420; A38420.
DR   HSSP; P04002; 1ATF.
KW   ANTIFREEZE PROTEIN; GLYCOPROTEIN; POLYPROTEIN; REPEAT;
KW   MULTIGENE FAMILY; SIGNAL.
FT   SIGNAL        1      ?       POTENTIAL.
FT   PROPEP        ?     37
FT   CHAIN        38    822       ANTIFREEZE GLYCOPEPTIDE POLYPROTEIN.
FT   PEPTIDE      38     51       AFGP8.
FT   PEPTIDE      55     68       AFGP8.
FT   PEPTIDE      72     85       AFGP8.
FT   PEPTIDE      89    102       AFGP8.
FT   PEPTIDE     106    119       AFGP8.
FT   PEPTIDE     123    136       AFGP8.
FT   PEPTIDE     140    153       AFGP8.
FT   PEPTIDE     157    170       AFGP8.
FT   PEPTIDE     174    187       AFGP8.
FT   PEPTIDE     191    204       AFGP8.
FT   PEPTIDE     208    221       AFGP8.
FT   PEPTIDE     225    238       AFGP8.
FT   PEPTIDE     242    255       AFGP8.
FT   PEPTIDE     259    272       AFGP8.
FT   PEPTIDE     276    289       AFGP8.
FT   PEPTIDE     293    306       AFGP8.
FT   PEPTIDE     310    323       AFGP8.
FT   PEPTIDE     327    340       AFGP8.
FT   PEPTIDE     344    357       AFGP8.
FT   PEPTIDE     361    374       AFGP8.
FT   PEPTIDE     378    391       AFGP8.
FT   PEPTIDE     395    408       AFGP8.
FT   PEPTIDE     412    425       AFGP8.
FT   PEPTIDE     429    442       AFGP8.
FT   PEPTIDE     446    459       AFGP8.
FT   PEPTIDE     463    476       AFGP8.
FT   PEPTIDE     480    493       AFGP8.
FT   PEPTIDE     497    510       AFGP8.
FT   PEPTIDE     514    527       AFGP8.
FT   PEPTIDE     531    544       AFGP8.
FT   PEPTIDE     548    561       AFGP8.
FT   PEPTIDE     565    578       AFGP8.
FT   PEPTIDE     582    595       AFGP8.
FT   PEPTIDE     599    612       AFGP8.
FT   PEPTIDE     616    629       AFGP8.
FT   PEPTIDE     633    646       AFGP8.
FT   PEPTIDE     650    663       AFGP8.
FT   PEPTIDE     667    680       AFGP8.
FT   PEPTIDE     684    697       AFGP8.
FT   PEPTIDE     701    714       AFGP8.
FT   PEPTIDE     718    731       AFGP8.
FT   PEPTIDE     735    748       AFGP8.
FT   PEPTIDE     752    768       AFGP7.
FT   PEPTIDE     772    788       AFGP7.
FT   PEPTIDE     792    805       AFGP8.
FT   PEPTIDE     809    822       AFGP8-LIKE.
SQ   SEQUENCE   822 AA;  74643 MW;  5C3109BA CRC32;

  Query Match             8.4%;  Score 369.5;  DB 1;  Length 822;
  Best Local Similarity   25.9%;  Pred. No. 1e-09;
  Matches  184;  Conservative   23;  Mismatches  314;  Indels  189;  Gaps   25;

Qy      82 CVHAALAGC-----TLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAP 136
           ||  ::  |       |   | |   |||        | |  | |      | |  | |
Db      21 CVCVSVCVCETAVTAAPAATAATAATPATAALNFAATAATPATPATPALIFAATAATAAT 80

Qy     137 TTTQA---PTTTQAPTTTQAPTTTTQAPTTTTQAPTT-------TTQAPTTTQAPTTTQA 186
             | |     |   | |   |     |   |   | |       |   | |  |    |
Db      81 PATAALNFAATAATPATAATPALIFAAAAATAATPATAALNFAATAATPATAATPALIFA 140

Qy     187 PTTTQAPTTITQA------ATTPA-ATTP----AATTPAATTPAATTPA---ATTPGVPA 232
             |   | |  | |       | ||| | ||    |||   | |||  |||   | |   ||
Db     141 ATAATAATPATPAFHFAATAATPATAATPALIFAATAATAATPA--TPAFHFAATAATPA 198

Qy     233 PTSAPVWPPICELLPNGCPA----DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYF 288
             : |           ||    :|      |          :   |          | |
Db     199 TAATPALIFAATAATAATPATAALNFAATAATPATAATPALIFAATAATAATPATAALNF 258

Qy     289 NPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRC 348
                 :||   |:   : | | |                 ||
Db     259 AATAATAATPATAACNFAATAATPATAAT--------------PA-------------- 290

Qy     349 DKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCP 408
                                        |:                    :
Db     290 --------------------------------LI--------------------FAA 294

Qy     409 TEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHF-----SPAT 463
           |     |   |  :|:     |  :    :      | | |    |:|     :|||
Db     295 TAATAATPATAAC--NFAATAATPATAATPALIFAATAATAATPATAALNFAATAATPAT 352

Qy     464 ------------QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAP 511
                      :  :| ||          |   : | |  | | |   | |||  | |
Db     353 AATPALIFAATAATAATPATAAL--------NFAATAATPATAATPALIFAATAATAATP 404

Qy     512 -----------STVVPPATPP---ATAAPVPPTTAIPTP--APTAAPTAAPTTAAPESPT 555
                    :|   ||||    |  |  | | | |    | |||  | | |||
Db     405 ATAALNFAATAATAATPATPAFNFAATAATPATAATPALIFAATAATAATPATAALNFAA 464

Qy     556 TVTVPPTAAPTA---APTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVT 612
           |   | |||  |   | | |       |  :| |||   | |||  | | ||      |
Db     465 TAATPATAATPALIFAATAA------TAATAATAALHFAATAATAATPATAALNFAATAA 518

Qy     613 SPPTA----------APTTAAPAP---NTTVTVPPTAAPTTAAPAPNTTVTAPPTAA-PT 658
           :| ||          | | | ||    |   |    | | |||     |   | ||| |
Db     519 TPATAATPALIFAATAATAATPATAAFNFAATAATAATPATAALNFAATAATPATAATPA 578

Qy     659 TAAPAPNTTVTVPPTA----APTAAPPTVAHAPNTTAAPVTTTSAPATTP 704
               |   |   | ||    | ||| | |  |    |    |:|   ||
Db     579 LIFAATAATAATPATAALNFAATAATPATAATPALIFAATAATAATPATP 628


RESULT   9
VTP3_TTV1V
ID   VTP3_TTV1V     STANDARD;      PRT;   474 AA.
AC   P19275;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE   VIRAL PROTEIN TPX.
OS   THERMOPROTEUS TENAX VIRUS 1 (STRAIN VT3) (TTV1).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; LIPOTHRIXVIRIDAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90245666.
RA   NEUMANN H., ZILLIG W.;
RL   NUCLEIC ACIDS RES. 18:2171-2171(1990).
DR   EMBL; X14717; E1192689; -.
DR   PIR; S06686; S06686.
DR   PIR; S15921; S15921.
KW   REPEAT.
FT   DOMAIN      278    467       3 THR-PRO REPEATS REGIONS AND TWO NEAR
FT                                IDENTICAL REPEATS.
FT   REPEAT      278    367       THR-PRO(N).
FT   REPEAT      368    377
FT   REPEAT      378    436       THR-PRO(N).
FT   REPEAT      437    446
FT   REPEAT      447    467       THR-PRO(N).
SQ   SEQUENCE   474 AA;  49507 MW;  BAC5D831 CRC32;

  Query Match             8.4%;  Score 367.5;  DB 1;  Length 474;
  Best Local Similarity   32.6%;  Pred. No. 8.5e-10;
```

```
  Matches  107;  Conservative   20;  Mismatches  132;  Indels   69;  Gaps   13;

Qy    394 EVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPC 453
          |:|: : |       |:|:      ||  |:  :            | |   |  |
Db    192 EIDISQFAAF-----TDPLVMYLSINGSTANVKV-----------YKQ---GSNI---- 228

Qy    454 PGNLHFSPATQSCESPVTAGCQ----------VFECDSDNQCTSTAAPTAAPTAAPTAAP 503
           |:  : :|    :|  ||            |   : | | | |   : :||  ||  |
Db    228 -GTVSGNYSTTPYGNPSMAGYGTVDKHYANFIVLPYEPDPQVTVTPISSPSPTPTPTPTP 286

Qy    504 TAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTA 563
          | ||  |:   |  ||  |  | |  |  ||| ||  ||  ||     :||     ||
Db    287 TPTPTPTPT---PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT-----PTP 338

Qy    564 APTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPT-TA 622
           ||  ||          |  ||  ||  ||  ||  ||  :  :   ||  ||  || |
Db    339 TPTPTPTP---------TPTPTPTPTPTPTPTPTPTPTYDITYVVFDVTPSPTPTPTPTP 389

Qy    623 APAPNTTVTVPPTAAPT-TAAPAPNTTVTAPPTAAPT-TAAPAPNTT---------VTVP 671
          | |  | |  ||  || |  | | | |  ||  || |  | |  |          ||
Db    390 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTYDITYVIFDVTPS 449

Qy    672 PTAAPTAAPPTVAHAPNTTAAPVTTTSA 699
          ||  ||  |      |  |  | :|||:
Db    450 PTPTPTPTP-----TPTPTPTPTSTTSS 472


RESULT  10
Y091_NPVOP
ID   Y091_NPVOP     STANDARD;      PRT;   279 AA.
AC   O10341;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 29.3 KD PROTEIN (ORF92).
OS   ORGYIA PSEUDOTSUGATA MULTICAPSID POLYHEDROSIS VIRUS (OPMNPV).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EUBACULOVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97271300.
RA   AHRENS C.H., RUSSELL R.R., FUNK C.J., EVANS J., HARWOOD S.,
RA   ROHRMANN G.F.;
RL   VIROLOGY 229:381-399(1997).
CC   -!- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV.
DR   EMBL; U75930; G1911338; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   279 AA;  29289 MW;  AABB6954 CRC32;


  Query Match             8.1%;  Score 357;  DB 1;  Length 279;
  Best Local Similarity   32.4%;  Pred. No. 1.6e-09;
  Matches  103;  Conservative   34;  Mismatches  111;  Indels   70;  Gaps   14;

Qy    484 QCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPT 543
          |  |  |  :||  || :||  || :|       ||  |  | |  |  |||:||  |
Db     26 QNIETKPPLPSPTPTPTPSPTPTPTPSP-------TP--TPTPTPTPTPTPSPTPTPA 76

Qy    544 AAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTTA 603
           :||             || :|| :||             |:  || :|| :|| :||
Db     77 LSPT------------PTPSPTLSPT------------PSPTPTPSPTPSPTPSPTP- 110

Qy    604 VPEIPTTVTSPPTAAPT-TAAPAPNTTVTVPPTAAPT-TAAPAPNTTVTAPPTAAPT-TA 660
                |  || :|| | :| |: | |  ||  |: | :| |: | |  || :|| |
Db    110 --------TPSPTPSPTPTPSPTPSPTPTPSPTPTPSPTPSPTPSPTPTPSPTPSPTPTP 161

Qy    661 APAPNTTVTVPPTAAPTAAP-PTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPIN 719
          :| |: | |  || :||  | ||   :|:    |:   |:  || :  |   || |
Db    162 SPTPSPTPTPSPTPSPTPPPSPTPPPSPSPLGDPMYFPSSVGTTDQLRDYIRPLCNGWPR 221

Qy    720 PCVEECNVLPWAHADCDKYWVCDGNNQVL--VVCSEGLQFNPTTKTCDFACNVGCVRSNI 777
          |  |   :||   || :   |   :  |  | |::   :: |          |||
Db    222 P--RETYAVPW---DCQRVMHCSYFSLPLHEVTCNDDSAYSFTAD--------GCV---- 265

Qy    778 QMSESYEGVQVFIPWNKL 795
            |:    | | | |
Db    265 ---PHYQSDCPFYPLNTL 279


RESULT  11
VG50_HSVI1
ID   VG50_HSVI1     STANDARD;      PRT;   670 AA.
AC   Q00130;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL GENE 50 PROTEIN.
GN   50.
OS   ICTALURID HERPESVIRUS 1 (CHANNEL CATFISH VIRUS) (CCV).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AUBURN 1;
RX   MEDLINE; 92087490.
RA   DAVISON A.J.;
RL   VIROLOGY 186:9-14(1992).
DR   EMBL; M75136; G331260; -.
DR   PIR; F36791; F36791.
KW   HYPOTHETICAL PROTEIN; REPEAT.
FT   REPEAT      143    158       1.
FT   REPEAT      171    186       2.
FT   REPEAT      200    214       3.
FT   REPEAT      215    233       4.
FT   REPEAT      234    252       5.
FT   REPEAT      253    268       6.
FT   REPEAT      279    293       7.
FT   REPEAT      294    309       8.
FT   REPEAT      320    334       9.
FT   REPEAT      335    349       10.
FT   REPEAT      362    376       11.
FT   REPEAT      377    391       12.
FT   REPEAT      392    406       13.
FT   REPEAT      407    421       14.
FT   REPEAT      422    436       15.
FT   REPEAT      437    452       16.
FT   REPEAT      464    477       17.
FT   REPEAT      478    493       18.
FT   REPEAT      504    517       19.
FT   REPEAT      518    531       20.
FT   REPEAT      532    545       21.
FT   REPEAT      546    559       22.
FT   REPEAT      560    573       23.
FT   REPEAT      574    587       24.
FT   REPEAT      588    601       25.
FT   REPEAT      602    615       26.
FT   REPEAT      616    629       27.
SQ   SEQUENCE   670 AA;  64174 MW;  FEC4F61D CRC32;


  Query Match             8.0%;  Score 352;  DB 1;  Length 670;
  Best Local Similarity   26.8%;  Pred. No. 5e-09;
  Matches  187;  Conservative   30;  Mismatches  252;  Indels  230;  Gaps   33;

Qy     96 PAETTQAPATTQA----------------PTTTQAPTT--TTQAPT------TTTQAPT 130
          | | :  | ||  :                  ||    |||   | ||       || :  |
Db     72 PREPSSLPRTTPSLAHTTISKMISLGTRPRPTIPGVPTTIPNTDAPVDPGSVHTTARVVT 131

Qy    131 --TTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPT 188
            || | |||   | |       |   || | |    |   || | |     |  ||  |
Db    132 DITTKQTPTT---PATPAGANDTANITTATPAGANDTANITTATPAGANDTANITTATPA 188

Qy    189 TTQAPTTITQAATTPAATTPAATTPAATTPAAT------TPAATTPGVP----------A 232
          | :|  : |||     | |   |||         |   ||| :|          |
Db    189 GAN-DTAVTTTSATPAGANDTAVTTTPATPAGANDTANDTVVTTTPAMPAGANDTANGTA 247

Qy    233 PTSAPVWPP-------ICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEG 285
          |: |  |         |    | |
```

```
Db    248 VTTTPAMPAGANDTANITTATPTG---------------------------------- 272

Qy    286 LYFNPYVQRCDSPANV-----ECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDW 340
                  :  |||       |  ::  | :  |  |   :|      |  ||   :
Db    272 ---------ANDTANVTMPAGATDTVVTTTPAMPTGAND-TANITTATPAG--ANDTANV 319

Qy    341 LLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDED 400
           :| |                                | :      :| | ::  :| :
Db    320 TMPAG------------------------------ATDTVVTTTPAMPAGANDTANVTKP 349

Qy    401 A-CTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHF 459
          |  |     | |     :| |                                 |:    |
Db    350 AGSTDTVVTTTP----AMPTGAT------------------------DTVVTTTP----- 377

Qy    460 SPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPAT 519
          :  | : :: ||                |: | || |     |  | | |  |  |||   |
Db    377 AMPTGATDTVVT-------------TTPAMPTGATDTVVTTTP-AKPAGANGTVV-TTT 420

Qy    520 P--PATAAPVPPTTAIPTPAPT--AAPTAAPTTAAPESPTTVTVPPTAA----------P 565
          |  || |     |||  |||     |    || |     || || || |          |
Db    421 PAMPAGANDTVVTTAPATPAGANDTANVTKPTGATDTVVTTATVKPTGATGTVTTTTAKP 480

Qy    566 TAAPTTAVPEIPITVTSAPTAAPTAAPTAA------PTAAPTTAVPEIPTTVTSPPTAA- 619
          | |  ||    |   |   |   || || |       || || |   : || |: || |
Db    481 TGANDTANVTKPTGATGTVTTT-TAKPTGATGTVTVATAKPTGATGTV-TTTTAKPTGAN 538

Qy    619 ---PTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPP-----TAAPTTAAPA-PNTTVT 669
             || |     | ||  | | | |    || || |     |   ||| ||  | |||
Db    539 GTVTTTTAKPTGATGTVTTTTAKPTGANGTVTTTTAKPAGANGTVTTTTAKPAGANGTVT 598

Qy    670 VPPTAAPTAAPPTVAHAPNTTAAP-------VTTTSAPA 701
             || |  |  ||     ||| |         |||: ||
Db    599 T-TTAKPAGANGTVT---TTTAKPAGANGTVTTTTAKPA 633


RESULT  12
EXTN_MAIZE
ID   EXTN_MAIZE     STANDARD;      PRT;   267 AA.
AC   P14918;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE   EXTENSIN PRECURSOR (PROLINE-RICH GLYCOPROTEIN).
GN   HRGP.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
OC   CYPERALES; GRAMINEAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. W64A, AND CV. E41;
RA   STIEFEL V., PEREZ-GRAU L., ALBERICIO F., GIRALT E., RUIZ-AVILA L.,
RA   LUDEVID M.D., PUIGDOMENECH P.;
RL   PLANT MOL. BIOL. 11:483-493(1988).
CC   -!- FUNCTION: STRUCTURAL COMPONENT IN PRIMARY CELL WALL.
CC   -!- TISSUE SPECIFICITY: MAINLY IN THE COLEOPTILE NODE AND ROOT TIP.
CC   -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC       SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC       GLYCOSYLATED.
DR   EMBL; X13499; G22508; -.
DR   EMBL; X13506; G22269; -.
DR   EMBL; M36912; G168455; -.
DR   EMBL; M36913; G168457; -.
DR   EMBL; M36914; G168459; -.
DR   PIR; S08314; S08314.
DR   MAIZEDB; 17152; -.
KW   REPEAT; CELL WALL; GLYCOPROTEIN; SIGNAL; STRUCTURAL PROTEIN;
KW   HYDROXYLATION.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?    267       EXTENSIN (PROLINE-RICH GLYCOPROTEIN).
FT   DOMAIN       18    253       HIGHLY REPETITIVE ZONE.
FT   REPEAT       18     33
FT   REPEAT       34     54
FT   REPEAT       55     70
FT   REPEAT       71     91
FT   REPEAT       92    107
FT   REPEAT      108    128
FT   REPEAT      129    144
FT   REPEAT      145    160
FT   REPEAT      161    179
FT   REPEAT      180    195
FT   REPEAT      196    211
FT   REPEAT      212    232
FT   REPEAT      233    253
FT   DOMAIN      261    265       EXTENSIN REPETIVE ELEMENT.
FT   VARÍANT     245    245       MISSING (IN STRAIN E41).
FT   VARIANT     261    261       S -> C (IN STRAIN E41).
SQ   SEQUENCE   267 AA;  28349 MW;  7CF5C68F CRC32;


  Query Match             7.7%;  Score 336.5;  DB 1;  Length 267;
  Best Local Similarity   33.6%;  Pred. No. 1.1e-08;
  Matches   91;  Conservative   18;  Mismatches  129;  Indels   33;  Gaps    9;

Qy    474 CQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPT--- 531
          |  |    :::  |   ||  |: |       ||  ||   |||: | |  | |||
Db      2 CPAFSIFFNSRRYSLTPPTYTPSPKPPTPKPTPPTYTPSP-KPPASKPPTPKPTPPTYTP 60

Qy    531 -TAIPTPAPTAAPTAAPTTAAP--ESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAA 587
           ||| ||  ||  |:   |  : ||    |||  |:  | |  |  | | | :|
Db     61 SPKPPTPKPT-PPTYTPSPKPPATKPPTPKPTPPTYTPSPKPPTPKP-TPPTYTPSPKPP 118

Qy    588 PTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNT 647
          |  ||  ||    |  |: ||   :|||  |:   | |  |   |||  |:   |   |
Db    119 ATKPPTPKPTPPTYTPSPKPPTPKPTPPTYTPSPKPPTPKPT---PPTYTPSPKPPTHPT 175

Qy    648 TVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNT-TAAP------------ 694
             |||  |:   | |  |   |||  |:  |||    | | | :|
Db    176 PKPTPPTYTPSPKPPTPKPT---PPTYTPSPKPPTPKPTPPTYTPSPKPPATKPPTPKPT 232

Qy    694 ----VTTTSAPATTPEDDDIDPPLPNDPINP 720
                |   ||| |      ||: : |  |
Db    233 PPTYTPTPKPPATKPPTYTPTPPVSHTPSPP 263


RESULT  13
MUA1_XENLA
ID   MUA1_XENLA     STANDARD;      PRT;   400 AA.
AC   P10667;
DT   01-JUL-1989 (REL. 11, CREATED)
DT   01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   INTEGUMENTARY MUCIN A.1 PRECURSOR (FIM-A.1) (PREPROSPASMOLYSIN).
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88227968.
RA   HOFFMANN W.;
RL   J. BIOL. CHEM. 263:7686-7690(1988).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE; 90316191.
RA   HAUSER F., GERTZEN E.M., HOFFMANN W.;
RL   EXP. CELL RES. 189:157-162(1990).
CC   -!- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC       INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AND STORED EXCLUSIVELY IN MATURE
CC       MUCOUS GLANDS OF THE SKIN.
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED. CONSIST OF ABOUT 70% CARBOHYDRATE
CC       AND 30% PROTEIN.
CC   -!- SIMILARITY: CONTAINS 4 P-TYPE (TREFOIL) DOMAINS.
DR   EMBL; M19971; G214792; -.
DR   PIR; A28172; A28172.
DR   HSSP; P01359; 1PCP.
DR   PROSITE; PS00025; P_TREFOIL; 3.
```

```
KW   REPEAT; SIGNAL; AMPHIBIAN SKIN; GLYCOPROTEIN.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    400       INTEGUMENTARY MUCIN A.1.
FT   DOMAIN       22     63       P-TYPE 1.
FT   DOMAIN       73    114       P-TYPE 2.
FT   DOMAIN      127    252       14 X 9 AA TANDEM REPEATS.
FT   DOMAIN      272    299       7 X 4 AA REPEATS OF E-T-T-T.
FT   DOMAIN      304    342       P-TYPE 3.
FT   DOMAIN      352    393       P-TYPE 4.
FT   DISULFID     23     49       BY SIMILARITY.
FT   DISULFID     33     48       BY SIMILARITY.
FT   DISULFID     43     60       BY SIMILARITY.
FT   DISULFID     74    100       BY SIMILARITY.
FT   DISULFID     84     99       BY SIMILARITY.
FT   DISULFID     94    111       BY SIMILARITY.
FT   DISULFID    312    327       BY SIMILARITY.
FT   DISULFID    322    339       BY SIMILARITY.
FT   DISULFID    353    379       BY SIMILARITY.
FT   DISULFID    363    378       BY SIMILARITY.
FT   DISULFID    373    390       BY SIMILARITY.
FT   CARBOHYD     63     63       POTENTIAL.
SQ   SEQUENCE   400 AA;  42641 MW;  6DEA5D96 CRC32;


  Query Match             7.5%;  Score 331;  DB 1;  Length 400;
  Best Local Similarity   30.1%;  Pred. No. 2.6e-08;
  Matches  114;  Conservative   17;  Mismatches  146;  Indels  102;  Gaps   18;

Qy    434 ESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAG-CQVFECDSD---------- 483
          |:|| : : |    :|             |: |    |||  :  ||  |
Db     40 EADC-RAVGCCFDSSILN----------TKWCFYNATAGPIKKLECSGDPTKRIDCGFP 87

Qy    483 ----NQC-----------------TSTAAPTAAP---TAAPTAAPTAAPTAAPSTVVPP 517
               ||                  |   | ||   ||: ||  |  ||  | |   |
Db     88 RITEKQCILRGCCFDSSISGVKWCYARTVITTPAPDTTTASTTAETTTVPT-TPETTTVP 146

Qy    518 ATPPATAAP-VPPTTAIP-TPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPE 575
          ||  |  |  | || :| ||  |  ||   ||  | :| | ||| |   |  |||  ||
Db    147 TTPETTTVPTTPETTTVPTTPETTTVPTTPETTTVPTTPETTTVPTTPETTTVPTT--PE 204

Qy    576 IPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPT 635
              |: ||   |   |  ||   || ||  | | |:  ||  ||    | |  |  ||
Db    205 ----TTTVPTTPET---TTVPTTPETTTVPTTPETTTASTTAETTTVPTTPET--TTEPT 255

Qy    636 AAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPV- 695
           |||    |   | |||  ||       || |   |   |    | | |   |  |
Db    256 TTPTTDTTPP----TLPPTPETTTETTTETTTETTTETTTETTTETTTAPPPECAADRVD 311

Qy    695 ------------------------------TTTSAPATTPEDDDIDPPLPNDPINPCV-- 723
                                         |   |||   |   :|| :  |   | :
Db    312 CGYSGITQADCEGKGCIFDSTIPETKWCFYTEAEAPARKAE-CTVDPSVRTDCGYPGITD 370

Qy    723 EECNVLPWAHADC--DKYW 739
          :||      : :|  |  |
Db    371 KECREKGCCYDECIPDVIW 389


RESULT  14
VGP3_EBV
ID   VGP3_EBV       STANDARD;      PRT;   907 AA.
AC   P03200; P03201;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   ENVELOPE GLYCOPROTEIN GP340 (MEMBRANE ANTIGEN) (MA) (CONTAINS:
DE   GYLCOPROTEIN GP220).
GN   BLLF1.
OS   EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).
OC   VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; GAMMAHERPESVIRINAE.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84270667.
RA   BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,
RA   GIBSON T.J., HATFULL G., HUDSON G.S., SATCHWELL S.C., SEGUIN C.,
RA   TUFFNELL P.S., BARRELL B.G.;
RL   NATURE 310:207-211(1984).
CC   -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
CC       B-CELLS.
CC   -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
CC       ENVELOPE.
DR   EMBL; V01555; G59109; -.
DR   EMBL; V01555; G832992; -.
DR   PIR; A03762; QQBE21.
DR   PIR; A03763; QQBE22.
DR   PIR; S33008; S33008.
KW   MEMBRANE; GLYCOPROTEIN; ANTIGEN; LATE PROTEIN; ALTERNATIVE SPLICING.
FT   VARSPLIC    502    698       MISSING (IN GP220).
FT   CARBOHYD     47     47       POTENTIAL.
FT   CARBOHYD     87     87       POTENTIAL.
FT   CARBOHYD    114    114       POTENTIAL.
FT   CARBOHYD    166    166       POTENTIAL.
FT   CARBOHYD    169    169       POTENTIAL.
FT   CARBOHYD    195    195       POTENTIAL.
FT   CARBOHYD    229    229       POTENTIAL.
FT   CARBOHYD    277    277       POTENTIAL.
FT   CARBOHYD    318    318       POTENTIAL.
FT   CARBOHYD    328    328       POTENTIAL.
FT   CARBOHYD    345    345       POTENTIAL.
FT   CARBOHYD    356    356       POTENTIAL.
FT   CARBOHYD    378    378       POTENTIAL.
FT   CARBOHYD    386    386       POTENTIAL.
FT   CARBOHYD    411    411       POTENTIAL.
FT   CARBOHYD    435    435       POTENTIAL.
FT   CARBOHYD    443    443       POTENTIAL.
FT   CARBOHYD    457    457       POTENTIAL.
FT   CARBOHYD    497    497       POTENTIAL.
FT   CARBOHYD    519    519       POTENTIAL.
FT   CARBOHYD    533    533       POTENTIAL.
FT   CARBOHYD    547    547       POTENTIAL.
FT   CARBOHYD    568    568       POTENTIAL.
FT   CARBOHYD    589    589       POTENTIAL.
FT   CARBOHYD    610    610       POTENTIAL.
FT   CARBOHYD    624    624       POTENTIAL.
FT   CARBOHYD    627    627       POTENTIAL.
FT   CARBOHYD    645    645       POTENTIAL.
FT   CARBOHYD    656    656       POTENTIAL.
FT   CARBOHYD    683    683       POTENTIAL.
FT   CARBOHYD    701    701       POTENTIAL.
FT   CARBOHYD    735    735       POTENTIAL.
FT   CARBOHYD    746    746       POTENTIAL.
FT   CARBOHYD    755    755       POTENTIAL.
FT   CARBOHYD    780    780       POTENTIAL.
FT   CARBOHYD    815    815       POTENTIAL.
FT   CARBOHYD    858    858       POTENTIAL.
FT   CARBOHYD    888    888       POTENTIAL.
SQ   SEQUENCE   907 AA;  94431 MW;  1EF4EBFD CRC32;


  Query Match             7.3%;  Score 321;  DB 1;  Length 907;
  Best Local Similarity   23.1%;  Pred. No. 1.3e-07;
  Matches  160;  Conservative   60;  Mismatches  192;  Indels  282;  Gaps   29;

Qy    100 TQAPATTQAPTTTQAPTTTQAPTTTQA---PTTTQAPTTTQAPTTTQAPTTTQAPTT 156
          | |  ||      ::|| :|| :||  |      | |||  |::|  ||   || :|  ||
Db    410 TNATTTTHKVIFSKAPESTTTSPTLNTTGFADPNTTTGLPSSTHVPTNLTAPAST-GPTV 468

Qy    157 TTQAPTTTTQAPTTTTQAPTT-----------TQAP--TTTQAPTTTQAP--TTITQAAT 201
              :|   |: | | ||:  :| |           ::||  |:: :| ||  |  |: | .| |
Db    469 STADVTSPTPAGTTSGASPVTPSPSPWDNGTESKAPDMTSSTSPVTTPTPNATSPTPAVT 528

Qy    202 TPAATTPAAT--TPAATTPAATTPAATTP--GVPAPTSAPVWPPICELLPNGCPADFDIH 257
          ||   || ||  ||| |||   || ||:|  |  :|||   |      ||
Db    529 TP---TPNATSPTPAVTTP---TPNATSPTLGKTSPTSAVTTP-----TPN--------- 569

Qy    258 LLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGN 317
                                                  ||      |: ||   ||  *
```

```
Db    569 ------------------------------------ATSPTL----GKTSPTSAVT--- 585

Qy    318 EDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQC 377
                                              | |
Db    585 ------------------------TPTPN------------------------------ 590

Qy    378 DHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDC 437
                 | |               |   ||  :   | ||
Db    590 -------ATSP-------------TLGKTSPTSAVT-TPTPNAT---------------- 613

Qy    438 GQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTA 497
                 | |:    |                        : :: |      :||   |:
Db    613 --------GPTVGETSP----------------------QANATNHTLGGTSPTPVVTS 641

Qy    498 APTAAPTAA-----------------PTAAPSTVVPPATPPATAAPVPPTTAIPT---- 536
            |  | :|                  |:: | |: |  :  :|:      |:| ||
Db    642 QPKNATSAVTTGQHNITSSSTSSMSLRPSSNPETLSPSTSDNSTSHMPLLTSAHPTGGEN 701

Qy    536 ---PAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVP-EIPITVTSAP-----TA 586
              | :  |   :|::| :|   |    : |  : |:  | |: :|  : |
Db    702 ITQVTPASISTHHVSTSSP-APRPGTTSQASGPGNSSTSTKPGEVNVTKGTPPQNATSPQ 760

Qy    587 APTAAPTAAPTAAPTTAVPEIPT-----------TVTSPPT---AAPTTAAPAPNTTVTV 632
            ||:   || ||   |        |           | | | |      ||  |  | |  :
Db    761 APSGQKTAVPTVTSTGGKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTPRPRYNATTYL 820

Qy    633 PPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAA 692
            ||:   |::   |  | |:||    |||      || ||||: |  :  ::      : |
Db    821 PPS---TSSKLRPRWTFTSPPV---TTA----QATVPVPPTSQPRFSNLSMLVLQWASLA 870

Qy    693 PVT----------------TTSAPATTPEDDDID 710
            :|                :||   |||  || :
Db    871 VLTLLLLLVMADCAFRRNLSTSHTYTTPPYDDAE 904


RESULT  15
CHI2_COCIM
ID   CHI2_COCIM     STANDARD;      PRT;   860 AA.
AC   P54197;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   ENDOCHITINASE 2 PRECURSOR (EC 3.2.1.14).
GN   CTS2.
OS   COCCIDIOIDES IMMITIS.
OC   EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C735;
RX   MEDLINE; 96144270.
RA   PISHKO E.J., KIRKLAND T.N., COLE G.T.;
RL   GENE 167:173-177(1995).
CC   -!- FUNCTION: MAY BE ASSOCIATED WITH ENDOSPORULATION.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC       N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC   -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC       HYDROLASES).
DR   EMBL; L41662; G1200192; -.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL;
KW   GLYCOPROTEIN.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    860       ENDOCHITINASE 2.
FT   CARBOHYD     90     90       POTENTIAL.
SQ   SEQUENCE   860 AA;  91395 MW;  00E9C2B0 CRC32;


  Query Match             7.3%;  Score 319.5;  DB 1;  Length 860;
  Best Local Similarity   23.1%;  Pred. No. 1.4e-07;
  Matches  148;  Conservative   72;  Mismatches  231;  Indels  191;  Gaps   22;

Qy     93 PGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTT----TTQAPTTTQAPTTTQAP 148
          | ||  |:   :|| | |:||  : :|   | |  | ||    :| :|::|   ||:

Db    342 PDPP--TSTVTSTTSASTSTQTSSQSTTMETKTLSASTTPSSPSTVSPSSTM-QTTSTGS 398

Qy    149 TTTQAPTTTTQAPTTTTQAPTTTTQAPTTT--QAPTTTQAPTTTQAPTTITQAATTPAAT 206
          |: :  || :| | :|| :  : :  | ||  | | :|   | : :  |:|  :  | :|
Db    399 TSIETVTTRSQEPPSTTISTRSASTEPVTTRSQEPPSTTISTRSASTETVTTRSQEPPST 458

Qy    207 TPAATTPAATTPAAT----TPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPH 262
          |   :| :|:|  :|    :|: |     |||
Db    459 T--ISTWSASTETSTSSQDSPSTTISTKSAPT---------------------------- 489

Qy    263 DKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDI 322
                                                |    :  |:  ::  : : :
Db    489 ----------------------------------GTVTTRSQDLPSTTISTRSPETET 512

Qy    323 DIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIEL 382
          :       | |:                    |  |   |                  |
Db    513 ETATTKSQGSPSI---------------------TLSTRSSSA-------------ET 536

Qy    383 VGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQ 442
          |        |  :            |||         |  :: |
Db    537 VSTRSQHSSSTTISTKS-------APTE--------TGTTSEHST--------------- 567

Qy    443 CVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAA 502
                      ::  |  : | |: :|   :    || :   || :|:   | : |
Db    567 ------------SMPVSTRSASTETVIT---RSQNSDSQSMTVSTRSPS---TESITTR 607

Qy    503 PTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPT 562
            :|:   ||   |    :|  |    | | |  | : | :||||         ||||
Db    608 SQGSPSETFSTKSVPVDTISTELPSQTPTTIITGTP-SDPVSAPTT---------TVPPN 657

Qy    563 AAPTAAPTTAVPEIPITVTS-APTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTT 621
            | ||:::  |   |:|:   |:  |  ||   |   |       | | :  ||  |
Db    658 PTLTLAPSSSTTEDRTTITTIITTSYVTVCPTGFTTVTITYTTTYCPETASLTPTQPPIP 717

Qy    622 AAPAP-----NTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVP-PTAA 675
            ||||      | ||| |  |||       ||||      || :|  |  | | |  |
Db    718 GAPAPPPDGWTTIVTVCPQCAPTP------TTVT---LTVPTRSAFLPARTETRPVVTVV 768

Qy    676 PTAAPPTVAHAPNTTAAPVT-TTSAPATTPEDDDIDPPLPND 716
          |    |     |: :   || ||:||||  :  : : |: :|
Db    769 PVPENPIKNVKPSESGDFVTVTTAAPATVTKTLEYNNPVDSD 810


Search completed: July 18, 1999, 01:41:36
Job time: 2319 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 02:42:01 ; Search time 32.5 Seconds
(without alignments)
1369.893 Million cell updates/sec

Title: US-09-294-663-4
Perfect score: 4395
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 807

Scoring table: BLOSUM62

Searched: 180763 seqs, 55169189 residues

Database :
SPTREMBL_8:*
1: sp_fungi:*
2: sp_human:*
3: sp_invertebrate:*
4: sp_mammal:*
5: sp_mhc:*
6: sp_organelle:*
7: sp_phage:*
8: sp_plant:*
9: sp_bacteria:*
10: sp_rodent:*
11: sp_virus:*
12: sp_vertebrate:*
13: sp_unclassified:*
14: sp_archea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4375	99.5	807	3	O18511	O18511 trichoplusi
2	4266.5	97.1	788	3	O18510	O18510 trichoplusi
3	542	12.3	1404	2	Q92954	Q92954 homo sapien
4	481	10.9	851	3	O17893	O17893 caenorhabdi
5	477.5	10.9	867	11	O39782	O39782 equine herp
6	473.5	10.8	3570	2	Q99552	Q99552 homo sapien
7	472.5	10.8	866	11	O39781	O39781 equine herp
8	459	10.4	801	3	Q23635	Q23635 caenorhabdi
9	453	10.3	770	3	Q20908	Q20908 caenorhabdi
10	431	9.8	2378	3	P91365	P91365 caenorhabdi
11	422.5	9.6	373	3	O76810	O76810 anopheles g
12	421.5	9.6	1188	8	Q41805	Q41805 zea mays (m
13	414.5	9.4	825	3	Q17921	Q17921 caenorhabdi
14	412.5	9.4	1611	1	O42854	O42854 schizosacch
15	405	9.2	507	12	O13028	O13028 boreogadus
16	403	9.2	622	2	Q14881	Q14881 homo sapien
17	398	9.1	750	11	O39307	O39307 equine herp
18	396.5	9.0	1777	3	Q22579	Q22579 caenorhabdi
19	390.5	8.9	1151	12	O57580	O57580 gallus gall
20	384	8.7	2187	10	P70670	P70670 mus musculu
21	380.5	8.7	846	3	O01699	O01699 caenorhabdi
22	373	8.5	1118	4	Q28226	Q28226 canis famil
23	370.5	8.4	412	3	O77242	O77242 heterodera
24	368	8.4	1131	3	Q18529	Q18529 caenorhabdi
25	365	8.3	353	3	Q25677	Q25677 podocoryne
26	348.5	7.9	279	2	Q14888	Q14888 homo sapien
27	347.5	7.9	350	8	Q41719	Q41719 zea diplope
28	346.5	7.9	5376	10	O88799	O88799 mus musculu
29	346	7.9	722	12	O13083	O13083 dissostichu
30	345.5	7.9	528	4	Q29071	Q29071 sus scrofa
31	345	7.8	232	3	Q27423	Q27423 drosophila
32	343.5	7.8	1229	3	Q94185	Q94185 caenorhabdi
33	342.5	7.8	2761	3	Q19522	Q19522 caenorhabdi
34	342.5	7.8	928	12	Q98906	Q98906 gallus gall
35	341	7.8	303	8	Q41814	Q41814 zea mays (m
36	340.5	7.7	216	4	Q28501	Q28501 macaca mula
37	340.5	7.7	328	8	Q42366	Q42366 zea mays (m
38	340	7.7	584	3	Q17802	Q17802 caenorhabdi
39	337	7.7	901	9	Q44562	Q44562 actinomyces
40	334.5	7.6	2225	3	O45881	O45881 caenorhabdi
41	325	7.4	907	11	Q66537	Q66537 epstein-bar
42	324.5	7.4	379	3	Q27929	Q27929 drosophila
43	322	7.3	3247	11	Q65553	Q65553 bovine herp
44	321	7.3	376	8	Q39353	Q39353 brassica na
45	319.5	7.3	946	8	O22015	O22015 cylindrothe

ALIGNMENTS

```
RESULT 1
O18511
ID   O18511      PRELIMINARY;      PRT;   807 AA.
AC   O18511;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   INSECT INTESTINAL MUCIN IIM22.
OS   TRICHOPLUSIA NI (CABBAGE LOOPER).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; LEPIDOPTERA; NOCTUOIDEA; NOCTUIDAE; TRICHOPLUSIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97341213.
RA   WANG P., GRANADOS R.R.;
RT   "Molecular cloning and sequencing of a novel invertebrate intestinal
RT   mucin cDNA.";
RL   J. BIOL. CHEM. 272:16663-16669(1997).
DR   EMBL; AF000606; G2224921; -.
SQ   SEQUENCE   807 AA;  83408 MW;  15E08C34 CRC32;

  Query Match            99.5%;  Score 4375;  DB 3;  Length 807;
  Best Local Similarity  99.4%;  Pred. No. 1.3e-119;
  Matches  802;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYG 60
          ||||||||||||||||||||||||||||||||| |||||  ||| ||||||||||||||
Db      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60

Qy     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180

Qy    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240

Qy    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300

Qy    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360

Qy    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420

Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 660

Qy    661 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 720
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    661 APAPNTTVTVPPTAAPTAAPPTVAPAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 720

Qy    721 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 780

Qy    781 ESYEGVQVFIPWNKLDEDIRQALNFEL 807
          |||||||||||||||||||||||||||
Db    781 ESYEGVQVFIPWNKLDEDIRQALNFEL 807
```

RESULT 2
O18510
ID O18510 PRELIMINARY; PRT; 788 AA.
AC O18510;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE INSECT INTESTINAL MUCIN IIM14.
OS TRICHOPLUSIA NI (CABBAGE LOOPER).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; LEPIDOPTERA; NOCTUOIDEA; NOCTUIDAE; TRICHOPLUSIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97341213.
RA WANG P., GRANADOS R.R.;
RT "Molecular cloning and sequencing of a novel invertebrate intestinal
RT mucin cDNA.";
RL J. BIOL. CHEM. 272:16663-16669(1997).
DR EMBL; AF000605; G2224919; -.
SQ SEQUENCE 788 AA; 81716 MW; 8F8ADE83 CRC32;

Query Match 97.1%; Score 4266.5; DB 3; Length 788;
Best Local Similarity 97.1%; Pred. No. 1.7e-116;
Matches 784; Conservative 0; Mismatches 4; Indels 19; Gaps 1;

```
Qy      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYG 60
          |||||||||||||||||||||||||||||||||| ||||| ||| ||||||||||||||
Db      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60

Qy     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180

Qy    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
```

```
Qy    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300

Qy    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360

Qy    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420

Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTV----------PPTAAPTTA 641

Qy    661 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    642 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 701

Qy    721 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    702 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 761

Qy    781 ESYEGVQVFIPWNKLDEDIRQALNFEL 807
          |||||||||||||||||||||||||||
Db    762 ESYEGVQVFIPWNKLDEDIRQALNFEL 788
```

RESULT 3
Q92954
ID Q92954 PRELIMINARY; PRT; 1404 AA.
AC Q92954;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGAKARYOCYTE STIMULATING FACTOR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA TURNER K.J., FITZ L.J., TEMPLE P., JACOBS K., LARSON D., LEARY A.C.,
RA KELLEHER K., GIANNOTTI J., CALVETTI J., FITZGERALD M., KRIZ M.J.,
RA FERENZ C., GROBHOLZ J., FRASER H., BEAN K., NORTON C.R., GESNER T.,
RA BHATIA S., KRIZ R., HEWICK R., CLARK S.C.;
RL BLOOD 78:279-279(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA MERBERG D.M., FITZ L.J., TEMPLE P., GIANNOTTI J., MURTHA P.,
RA FITZGERALD M., SCALTRETO J., KELLEHER K., PREISSNER K., KRIZ R.,
RA JACOBS K., TURNER K.;
RL (IN) PREISSNER K.T., ROSENBLATT S., KOST C., WEGERHOFF J.,
RL MOSHER D.F. (EDS.); BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS.:45-52;
RL ELSEVIER SCIENCE PUBLISHERS B.V. (1993).
RN [3]
RP SEQUENCE FROM N.A.
RA TURNER K.J., FITZ L.J., TEMPLE P., JACOBS K., LARSON D., LEARY A.C.,
RA KELLEHER K., GIANNOTTI J., CALVETTI J., FITZGERALD M., KRIZ M.J.,
RA FERENZ C., GROBHOLZ J., FRASER H., BEAN K., NORTON C.R., GESNER T.,
RA BHATIA S., KRIZ R., HEWICK R., CLARK S.C.;

```
RL   SUBMITTED (SEP-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U70136; G1572721; -.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR   PFAM; PF00045; hemopexin; 2.
DR   PFAM; PF01033; Somatomedin_B; 2.
SQ   SEQUENCE   1404 AA;  151090 MW;  6C1F4DEA CRC32;


  Query Match             12.3%;  Score 542;  DB 2;  Length 1404;
  Best Local Similarity   30.4%;  Pred. No. 1.6e-09;
  Matches  215;  Conservative   23;  Mismatches  194;  Indels  276;  Gaps   37;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT---QAPTTTTQAPTTTTQ-APTTTQ-AP 142
           | | || ||: || |      ||| : | :||     |||   ||||  : |||||: ||
Db     331 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 390

Qy     143 TTTQ--APTTTQAPTTTT---QAPTTTTQ-APTTTTQAPTTTQ--APTTTQ--APTTTQA 192
           || :  |||||: |  ||    ||||| : ||||| ||||  :  ||||  :  ||||  :
Db     391 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 450

Qy     193 PTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPA 252
           |     | |||   ||  |||    |    || |||  |||| ||  |
Db     451 P-----APTTPKEPTP--TTPKEPAPTTKEPAPTTPKEPAPT-APKKPA---------- 492

Qy     253 DFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPP 312
                                                |   :  :|   :      |||
Db     492 ------------------------------------PTTPKEPAPTTPK-----EPAPT 509

Qy     313 VTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSF 372
           |:                 |:
Db     510 TTKE----------------PS---------------------------------- 516

Qy     373 ELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLP---NGCPADFSIDH 429
                                          || |   || |      |
Db     516 ----------------------------------PTTP--KEPAPTTTKSAPT------ 533

Qy     430 LLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTA 489
                         |   | |     :| |    || |              |
Db     533 -----------------TTKEPAPTTTKSAPTTPKEPSPTT-------------TKEP 560

Qy     490 APTA----APTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
           |||     |||     |||     ||:|   || |  |  | |  ||   ||||     |
Db     561 APTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPA-PTTTKKPAP--TAPKEPAPTTPKETA 617

Qy     546 PTTAAPESPTT-----VTVPPTAAPTA----APTTAVPEIPITVT---SAPTAAPTAAPT 593
           |||    :|||      | |   |||     ||||  || |   |    |||    |||
Db     618 PTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTT--PEEPTPTTPEEPAPTTPKAAAPN 675

Qy     594 AAPTAAPTTAVPEIPTT-----VTSPPTAAPTT---------AAPAPNT---------- 629
                ||||     |||       |:|  ||||           ||| |
Db     676 TPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAP 735

Qy     629 TVTVPPTAAPTTAAPAP----NTTVTAPPTAAPTT-AAPAPNTTVTVPPTAAPTAAPPTV 683
           | |  ||:  |:  |||     |  | |   ||||   ||| |     ||     || |
Db     736 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP 794

Qy     684 AH------APNTTAAPVTTTS---APATTPEDDDIDP--PLPNDPINP 720
                   || ||  | :|||   || | |      |  | | |  |
Db     795 KKPAPKELAPTTTKGPTSTTSDKPAPTTPKETAPTTPKEPAPTTPKKP 842


RESULT   4
O17893
ID   O17893      PRELIMINARY;      PRT;   851 AA.
AC   O17893;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   F55B11.3.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA   BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA   DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA   HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA   KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA   MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA   RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA   SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA   VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA   WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z83318; E1186497; -.
SQ   SEQUENCE   851 AA;  92127 MW;  87CF937F CRC32;


  Query Match             10.9%;  Score 481;  DB 3;  Length 851;
  Best Local Similarity   25.0%;  Pred. No. 6.3e-08;
  Matches  160;  Conservative   22;  Mismatches  150;  Indels  308;  Gaps   16;

Qy      91 TLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTT 150
           | |  | : |  |:   : ||  |||||  |||||  |||||  |||    ||| : ||
Db     320 TRPTIPTKLTTTPSEL-STTTASVPTTTTSVPTTTTTVPTTTTTVPTT----TTTVSTTT 374

Qy     151 TQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAA 210
           |  |||||  |||||  ||||  |||    |||   |||  ||| |         | |
Db     375 TTVPTTTTTVPTTTTTVSTTTTTVPTT----TTTVPTTTTTVPTTTT--------TVPTT 422

Qy     211 TTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFY 270
           ||   ||   ||   ||  ||  |:
Db     423 TTSVPTT--TTTVPTTTTTVPTTTT---------------------------------- 446

Qy     271 QCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDN 330

Db     446 ------------------------------------------------------------ 446

Qy     331 GCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGG 390

Db     446 ------------------------------------------------------------ 446

Qy     391 ESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA 450
                                                                    |::
Db     446 ------------------------------------------------------TVS 448

Qy     451 RPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAA 510
                                                |:|  ||   | : |      |
Db     449 ---------------------------------TTTTTVPTTTTTVSTTT------TTV 468

Qy     511 PSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPT 570
           |:|     | | |   || ||   |  ||   |: |||      || ||| |    :  |
Db     469 PTTT---TTVPTTTTTVPTTT---TTVPTTT-TSVPTTTTTVPTTTTVPTTTTTVSTTT 521

Qy     571 TAVPEIPITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTV 630
           | |   | | |: ||   | : |       || ||   |||::  |  |||   || ||
Db     522 TTV---PTTTTTVPTTTTTVSTTTTTVPTTTTTVPTTTTTVSTTTTTVPTTTTTAPTTTT 578

Qy     631 TVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTT 690
           ||| |   ||  |   |||:   |  |||    | || ||| |                ||
Db     579 TVPTT---TTTVPTTTTTVSTTTTTVPTTTTTVPTTTTTVPTT---------------TT 620

Qy     691 AAPVTTT------SAPATTPEDDDIDPPLPNDPINPCVEE 724
             | |||      : | |   |  ::  : |   : | ||
Db     621 TVPTTTTPVLFVYTEPTTDYYDVLLEASMEN---STCGEE 657


RESULT   5
O39782
ID   O39782      PRELIMINARY;      PRT;   867 AA.
AC   O39782;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
```

```
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   MEMBRANE GLYCOPROTEIN.
OS   EQUINE HERPESVIRUS 1.
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC   ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BK343;
RA   KIRISAWA R., KOBAYASHI T., KAWAKAMI Y., IWAI H.;
RL   J. EQUINE SCI. 7:79-87(1996).
DR   EMBL; D88734; D1020846; -.
KW   MEMBRANE.
SQ   SEQUENCE   867 AA;  86630 MW;  DC3E8AA2 CRC32;


  Query Match             10.9%;  Score 477.5;  DB 11;  Length 867;
  Best Local Similarity   27.0%;  Pred. No. 8e-08;
  Matches  173;  Conservative   34;  Mismatches  152;  Indels  281;  Gaps   16;

Qy      78 SAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTT--TTQA 135
           | |:   || :  : |   : ||  | :|   ||| || :|  ||||| |||| || | 
Db      77 STQSSSTAATSS-SAPSTASSTTSIPTSTSTETTTTTPTASTTTPTTTTAAPTTAATTTA 135

Qy     136 PTT---TQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTT-TQAPTTTQAPT--TTQAPTT 189
           ||   | | |||    | |  ||||   |||| | ||  | | |||   |  || | ||
Db     136 VTTAASTSAETTTATATATSTPTTTTPTSTTTTTATTTVPTTASTTTDTTTAATTTAATT 195

Qy     190 TQAPTT--ITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLP 247
           | | ||   | |||| |||| |||| |||| |||| ||||    | |:|
Db     196 TAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATT--TAATTTA---------- 243

Qy     248 NGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEI 307

Db     243 ------------------------------------------------------------ 243

Qy     308 SPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAG 367

Db     243 ------------------------------------------------------------ 243

Qy     368 THFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSI 427

Db     243 ------------------------------------------------------------ 243

Qy     428 DHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTS 487
                                                                       |:
Db     243 -----------------------------------------------------------ATT 245

Qy     488 TAAPTAAPT--AAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAA 545
           ||| | | |  :| ||| | | |   :|     |. ||     |||  | | |   |||
Db     246 TAATTTAATTSSATTAATTTAATTTAATTTAATTTAAT------TTAATTTAATT--TAA 297

Qy     546 PTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTTAVP 605
           ||||  :  | |    |  ||| |||      | | :||:  |:  | | |: |: : 
Db     298 TTTAATTTAATTT---AATTTAATTTAATTTAATTTGSPTSGSTST-TGASTSTPSASTA 353

Qy     606 EIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAP--APNTTVTAPPTAAPTTAAPA 663
             |  ::  :|| ||: | | :  |   :|  || ||   | | | | |   | :|
Db     354 TSATPTSTSTSAAATTSTPTPTSAAT---SAESTTEAPTSTPTTDTTTPSEATTATTSP- 410

Qy     664 PNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATT 703
           :|||:   |:| | |  | :|    ::   |:|: |::|
Db     410 ESTTVSASTTSATTTAFTTESHTSPDSSTGSTSTAEPSST 449


RESULT   6
Q99552
ID   Q99552      PRELIMINARY;      PRT;  3570 AA.
AC   Q99552;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MUCIN MUC5B (FRAGMENT).
GN   MUC5B.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 97166151.
RA   DESSEYN J.L., GUYONNET-DUPERAT V., PORCHET N., AUBERT J.P., LAINE A.;
RT   "Human mucin gene MUC5B, the 10.7-kb large central exon encodes
RT   various alternate subdomains resulting in a super-repeat. Structural
RT   evidence for a 11p15.5 gene family.";
RL   J. BIOL. CHEM. 272:3168-3178(1997).
DR   EMBL; Z72496; E244318; -.
FT   NON_TER       1      1
FT   NON_TER    3570   3570
SQ   SEQUENCE   3570 AA;  361016 MW;  452C733D CRC32;


  Query Match             10.8%;  Score 473.5;  DB 2;  Length 3570;
  Best Local Similarity   22.5%;  Pred. No. 2.7e-07;
  Matches  223;  Conservative   85;  Mismatches  296;  Indels  385;  Gaps   41;

Qy      12 GLVAARPEVSDAEKNPAL--HEPHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRDC 69
           |  |  | ||    | :   :  |   |       |       |         | |
Db    2362 GPPAGTPHVSTTATTPTVTSSKATPFSSPGTATALPALRSTAT----TPTATSFTAIPSS 2417

Qy      70 APGTEFKFSAQTCVHAALAGCTLPGPPAET--TQAPATTQAPT---------------- 111
           : || :   :||    |       |    ||  |    || | |
Db    2418 SLGTTWTRLSQTTTPMATMSTATPSSTPETVHTSTVLTTTATTTGATGSVATPSSTPGTA 2477

Qy     111 -TTQAPTTTTQAPTTTTQAPTTTTQAP----TTTQAPTT-------TQAPTTTQAP---T 155
            ||: |||||   | |  :   | : |    :||  |||       :  | ||  |   |
Db    2478 HTTKVPTTTTTGFTVTPSSSPGTARTPPVWISTTTTPTTSGSTVTPSSIPGTTHTPTVLT 2537

Qy     156 TTTQ--------APTTTTQA----PTTTTQAPTTTQAPTTT---QAPTTTQAPTTITQAA 200
           ||||         |:::||     |:  | | | |   :||     | ||  |  :| |
Db    2538 TTTQPVATGSMATPSSSTQTSGTPPSLITTATTITATGSTTNPSSTPGTTPIPPELTTTA 2597

Qy     201 TTPAATTPAAT---------------------------------------------------- 212
           ||||||:   |
Db    2598 TTPAATSSTVTPSSALGTTHTPPVPNTTATTHGRSLSPSSPHTVRTAWTSATSGTLGTTH 2657

Qy     212 ----------TPAATTPAATT-------------------PAATTPG------------ 230
                    ||||||   ||                    |  ||||
Db    2658 ITEPSTGTSHTPAATTGTTTTSTPALSSPHPSSRTTESPPSPGTTTPGHTTATSRTTATA 2717

Qy     230 VPAPT-------SAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCS----NGYTF 278
           |: |       | |   ||  ::  ||                          ||:     |::
Db    2718 TPSKTRTSTLLPSQPTSAPITTVVTTGCEP-----------------QCAWSEWLDYSY 2759

Qy     279 EQRCPEGLYFNPYVQ-------RCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNG 331
           | |  |: |           |: |  :||      |  |: |        :| :::
Db    2760 PMPGPSGGDFDTYSNIRAAGGAVCEQPLGLECRATAQPGVPLGE--------LGQVVE-- 2810

Qy     332 CPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGE 391
           |  :| :  :  :  :  |:  |               |::|::       | |
Db    2810 CSLDFGL--VCRNREQVGKFKMC--------------FNYEIR-------VFC------ 2840

Qy     392 SEEVDVDEDACTGWYCPTEP-IEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA 450
                 |   :||: |      :|:  |       :|   :
Db    2840 ----------CNYGHCPSTPATSSTAMPSSTP---GTTWILTELTTTATTTASTGSTATP 2886

Qy     451 RPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAP----------TAAPTAAP 499
              ||    :|  :   || |         : :: ||:::|          ||  : |
Db    2887 SSTPGT---APPPKVLTSPATT-----PTATSSKATSSSSPRTATTLPVLTSTATKSTAT 2938

Qy     500 TAAP--------------TAAPTAAPSTVVPPATP-------------------PATA 524
           :  |               |  | |  ||: | :||                    :|
Db    2939 SVTPIPSSTLGTTGTLPEQTTTPVATMSTIHPSSTPETTHTSTVLTTKATTRATSSTSTP 2998

Qy     525 APVPPTTAI----------------PTPAPTAAP----------TAAPTTAA--------- 551
           :  | || |               ||  |:: |          | | |||:
```

```
Db   2999 SSTPGTTWILTELTTAATTTAGTGPTATPSSTPGTTWILTELTTTATTTASTGSTATLSS 3058

Qy    551 ----------PESPTTVTVPPTAAPTAAPTTA---VPEIPITVTSAPTAAPTAAPTAAPT 597
                  | :  ||| || :  ||: | |    | :  | |:    :  | |:::|
Db   3059 TPGTTWILTEPSTTATVTAPPGSTATASSTQATAGTPHVSTTATTPTVTSSKATPSSSPG 3118

Qy    598 AAPTTAVPEIPTTVTSP------------------------PTAAPTTAAPA--PNT-- 629
          |  ||:| : :| |:|                          |||  :|| |:  | |
Db   3119 TA--TALPALRSTATTPTATSFTAIPSSSLGTTWTRLSQTTTPTATMSTATPSSTPETVH 3176

Qy    629 TVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAP- 688
          | ||  | | || |     | :: |  | ||  |   || |   || |:::| |   |
Db   3177 TSTVLTTTATTTGATGSVATPSSTPGTAHTTKVP---TTTTTGFTATPSSSPGTALTPPV 3233

Qy    688 --NTTAAPVTTT----------SAPATT 703
            :||  | |||          | | ||
Db   3234 WISTTTTPTTTTPTTSGSTVTPSSIPGTT 3262


RESULT   7
O39781
ID   O39781      PRELIMINARY;      PRT;   866 AA.
AC   O39781;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   MEMBRANE GLYCOPROTEIN.
OS   EQUINE HERPESVIRUS 1.
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC   ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HH1;
RA   KIRISAWA R., KOBAYASHI T., KAWAKAMI Y., IWAI H.;
RL   J. EQUINE SCI. 7:79-87(1996).
DR   EMBL; D88733; D1020845; -.
KW   MEMBRANE.
SQ   SEQUENCE   866 AA;  86463 MW;  7200333A CRC32;


  Query Match             10.8%;  Score 472.5;  DB 11;  Length 866;
  Best Local Similarity   25.5%;  Pred. No. 1.1e-07;
  Matches  179;  Conservative   41;  Mismatches  175;  Indels  307;  Gaps   18;

Qy     78 SAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTQAPTTTQAPTT--TTQA 135
          | |:   || :  : |   : ||  | :|  ||| || :|  ||||| ||||  || |
Db     77 STQSSSTAATSS-SAPSTASSTTSIPTSTSTETTTTTPTASTTTPTTTTAAPTTAATTTA 135

Qy    136 PTT-------TQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTT-TTQAP 187
            ||       | | |||   | |  |||||   ||||   | ||  |||    | || |
Db    136 VTTAASTAASTSAETTTATATATSTPTTTTPTSTTTT---TATTTVPTTASTTTDTTTAA 192

Qy    188 TTTQAPTT--ITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICEL 245
          |||  | ||   | |||| |||| |||| |||| |||| ||||    | |:|
Db    193 TTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATT--TAATTTA--------- 242

Qy    246 LPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDG 305

Db    242 ------------------------------------------------------------ 242

Qy    306 EISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCG 365

Db    242 ------------------------------------------------------------ 242

Qy    366 AGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADF 425

Db    242 ------------------------------------------------------------ 242

Qy    426 SIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQC 485
                                                 | |  :  |         :
Db    242 ---------------------------------ATTSSATTAATTTAATTTAATTTAA 266

Qy    486 TSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPT--AAPT 543
          |:||| | | |   ||| | | |   :|     |  ||      | |  | | |  |  |
Db    267 TTTAATTTAATT--TAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 324

Qy    544 AAPTTAAPESPTTVTVPPTAAPTAAP-----TTAVPEIPITVTSAPTAAPTAAPTAAPTA 598
          || || :| | :| |   | | |: |     |:| |    |  :| |: ||  ||:| |:
Db    325 AATTTGSPTSGSTST---TGASTSTPSASTATSATPTSTSTSAAATTSTPT--PTSAATS 379

Qy    599 APTT--AVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAA 656
          | :|  |    ||| |: |: | |      :|||:   |:| |||    :   |:| ::
Db    380 AESTTEAPTSTPTTDTTTPSEATTATTSPESTTVSASTTSATTTAFTTESH--TSPDSST 437

Qy    657 PTTAAPAPNTTVTVPPTAA----------------------PTAAPPTVAHAPNTTAA 692
          :|:   |::| |: |: |                        ||    ::  : :|| |
Db    438 GSTSTAEPSSTFTLTPSTATPSTDQFTGSSASTESDSTDSSTVPTTGTESITESSSTTEA 497

Qy    693 ----------------PVTTTSAPATTPEDDDIDPPLPNDP 717
                          |   |::  |||       | |  |
Db    498 STNLGSSTYESTEALETPDGNTTSGNTTPS------PSPRTP 533


RESULT   8
Q23635
ID   Q23635      PRELIMINARY;      PRT;   801 AA.
AC   Q23635;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO LONG TANDEM REPEAT REGION OF SIALIDASE.
GN   ZK84.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA   BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA   DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA   HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA   KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA   MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA   RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA   SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA   VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA   WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   KIRSTEN J.;
RL   SUBMITTED (APR-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U23181; G726436; -.
SQ   SEQUENCE   801 AA;  77123 MW;  7288A89D CRC32;


  Query Match             10.4%;  Score 459;  DB 3;  Length 801;
  Best Local Similarity   26.7%;  Pred. No. 2.6e-07;
  Matches  197;  Conservative   42;  Mismatches  295;  Indels  204;  Gaps   28;

Qy     10 ALGLVAARPE--VSDA-EKNPALHEPHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAP 66
          | | | :  |    || |  | : |     | ||:          |     |       |
Db    194 ASGYVTSEQEQGAGDAPEPAPVVEETPAPTPAAEE----------TPAPATEASNAVGTP 243

Qy     67 RDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTT-TTQAPTTT 125
              ||    :| : |          | | || | ||||:       |||   | || |
Db    244 EGYVDGTA---NAASAVAEPAVETPAPAPAAEETPAPATSSEEAPAPAPTAEETPAPETV 300

Qy    126 TQAP------------TTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQ 173
          : ||              |  ||::  |       :   | ||     |::::    :|:
Db    301 SAAPEAANSYDSAGGDAATAPAPSSEAADAAAPTDSAASADTTAALVDTSSSEHAAESTE 360
```

```
Qy    174 APTTTQAPT-TTQAPTTT----------QAPTTITQAATTPAA-TTPAATTPAATTPAAT 221
          || |  | | || ||:            :|::| :    |||  |||  : ||    |
Db    361 APATDIAATETTPAPSVPAPVADAAAGYDSPSSIPEETPAPAAEDTPAPASAAAEETPAP 420

Qy    222 TPAATTPGVPAPTSAPVWPPICELLPNGCPADF----DIHLLIPHDKYCNLFYQCSNGYT 277
           |||     ||| :|   |   :      |||     |:    |               |
Db    421 APAAEE--TPAPETASAAP---DAAGGAAPADVAAPADVATTAPE-------------T 461

Qy    278 FEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPV---TEGNEDEDIDIGDLLDNGCPA 334
          :   | |  |        ||:     |:: || |   ||   |   ||        ||
Db    462 SSAQSAAGSYDVP-----SEPAS-----EVT-APIVESATEAPSDSAAPIGPAASEPAPA 510

Qy    335 NFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEE 394
                                                |   |   |     |  |
Db    511 ------------------------------------PIEAPATDAATLETAPAPAAEPAP 534

Qy    395 VDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCP 454
          :|  |:  |:   |  | |
Db    535 A---AEAAAGYDAPSSVPEETPAP---------------------------------- 556

Qy    455 GNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTV 514
          :||    :|                 || |  | ||  |    |  ||:
Db    556 -----APAADETPAP-------------------APAAEETPAPAPAAEETPAPAPAAE 590

Qy    515 VPPATPPA---TAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTT 571
            ||| ||   | || |     | ||| |  | ||  |  |:|      |    | ||
Db    591 ETPATAPAAEETPAPAPAADETPAPAPAAEETPAPAPAVEETPAPA---PAVEETPAPAP 647

Qy    572 AVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVT 631
          |  | |     || ||  ||| |  | ||  |  | |    :|  ||  | ||||
Db    648 AAEETP---APAPAAAEYAAPVAEETPAPAPAAEETP----APAQAAEETPAPAP----- 696

Qy    632 VPPTAAPTTAAPAPNTTVT-APPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTT 690
            ||  | | ||    | ||  ||   ||||    |  | |  |||  :  :
Db    696 ----AAEETPASAPAAEETPAPAPAAEEAPAPAPAAPETPAPAAEGAAAPAPIVSSGYDA 751

Qy    691 AA---PVTTTSAPATTPE 705
          |     |  :||||||: :
Db    752 AQSTDTVVASSAPATSSD 769
```

```
RESULT   9
Q20908
ID   Q20908      PRELIMINARY;      PRT;   770 AA.
AC   Q20908;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   F56H9.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BURTON J.;
RL   SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA   BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA   DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA   HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA   KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA   MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA   RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA   SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA   VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA   WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
```

```
DR   EMBL; Z74473; E249616; -.
SQ   SEQUENCE   770 AA;  81625 MW;  771AB928 CRC32;


  Query Match             10.3%;  Score 453;  DB 3;  Length 770;
  Best Local Similarity   23.7%;  Pred. No. 3.7e-07;
  Matches  185;  Conservative   53;  Mismatches  314;  Indels  228;  Gaps   22;

Qy     71 PGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPT 130
          | |    :  | | :     |       ||  | ||   |||  | |||   ||||  ||
Db     55 PSTTTSSATTTTVPSTTTSST-----TTTTVPPTTTSTTTTTTVPPTTTSTTTTTTVPPT 109

Qy    131 TTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTT 190
          ||:   |||  |||| : |||  | ||| : ||||  |||||   |||  ||||   |||
Db    110 TTSTTTTTTVPPTTTSSTTTTTVPPTTTSSTTTTTVPPTTTTSTTTTTVPPTTTSTTTTT 169

Qy    191 QAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGC 250
            | | |  |||       :||   | |  ||   ||  ||  |::         ||
Db    170 TVPATTTSTATTTTVPPTTSTTTTTVPPTTTSTTTTTTVPPTTTSSTTTTTTTTLPTTT 229

Qy    251 PADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPA 310
                ::||    |:|     :  ||      :|  | | |
Db    230 TT-----MMIP---ICSL-----DSLTF----GKGDNFQPQV----------------- 255

Qy    311 PPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRC--------------------D 349
                | |:   :|: | |   ::  |     |
Db    255 --------DIDVTYSNLVANIIPGTMDMQSLSTMQISCTAIAGYNAYMIFDVFNQPAENQ 306

Qy    350 KYYQCVHGNLVERR----------------------CGAGTHFSFEL------------- 375
            : | |  |    |                      | : |  :| |
Db    307 GFPQSVTVNAQCNRILINLFMFFKFLSYEKTSFYILCTSKTLSNFILLTVYFLYIGPTEL 366

Qy    375 --QQCDHIEL---------VGCTLPGGESEEVDVDEDACTGWYCPTE------------- 411
              |   : |         :|  | |  :: :          |: ||
Db    367 LYTQIGSMNLNTYLNQTMGLGMYLQGPITQMIITINRFLVIWFTPTHVPQYSHRITLGAL 426

Qy    411 PIEWEP---------LPN--------------GCPADFSIDHLLPHESDCGQYLQCV--- 445
            : |           ||              |      :|:  : : |   :  :
Db    427 SVSWITVTWLSTLIGLPGFDNNNNFDIIINLANCRVPIGFEHIGYYSTPCNNQITIIVVS 486

Qy    445 ---------------------------------HGQTIARPCPGNLHFSPATQSCESPV 470
                                            : : : :     ||    |
Db    487 GIFLIGFLTNFMNFMIGGKLIYTWKTLTIFRFLSNFKCVMKWISTLLHILLLT------- 540

Qy    471 TAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPT 530
              :: |  | | :| :||| ||   | | |       :  |   | | |    | |
Db    540 ------YKLDVINSCIAT-SPTADPTTTTTEATTTTEITTTTEEVTTTTEPVTTTTTPTT 592

Qy    531 TAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTA 590
          |:  :   |  ||   ||    :||| |    :   |  |||       | |:  || ||
Db    593 TSTTSTTTTTPTTTTTT----TPTTTTSTTSTTTTTPTTTTSTTTTTTTTTTTTATPTT 648

Qy    591 APTAAPTAAPTTAV--------PEIPTTVTSP---PTAAPTTAAPAPNTTVTVPPTAAPT 639
            |  |   |             |:|   ||      |  | |       |||
Db    649 TTTTMPPCNPNAVSLGMGDANNPQIFIDVTYNNYLSTTDPVTGIVTSTMTVTCSALNGYN 708

Qy    640 T--------AAPAPNTTV--TAPPTAAPTTAAPAPNTTVTV-PPTAAPTAAPPTVAHAPN 688
          |          || |  :  |   |   |:      |  ||:   |     :  |   |||
Db    709 TYMLFNGGQGGPADNQNLPQTISITLVCTSDTMVWNYEVTINGVTYTRAVSSVTCQQAPN 768
```

```
RESULT  10
P91365
ID   P91365      PRELIMINARY;      PRT;  2378 AA.
AC   P91365;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   CODED FOR BY C. ELEGANS CDNA YK50C7.5.
GN   K06A9.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA   BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA   DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA   HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA   KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA   MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA   RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA   SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA   VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA   WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   GEISEL C., GATTUNG S.;
RL   SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U80846; G1707247; -.
SQ   SEQUENCE   2378 AA;  225242 MW;  765EDB23 CRC32;

  Query Match             9.8%;  Score 431;  DB 3;  Length 2378;
  Best Local Similarity   25.1%;  Pred. No. 3.4e-06;
  Matches  158;  Conservative   50;  Mismatches  144;  Indels  278;  Gaps   26;

Qy    111 TTQAPTTTTQAPTTTTQAPTTTTQ-----APTTTQAPTTTQAPTTTQAPTTTTQAPTTTT 165
Db    223 TTLGPTTTTITPTTTVVPASTSTQPAPASSPTTQPAPAT--SPTTQPAPATS----PTTQ 276

Qy    166 QAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPA- 225
Db    277 PAPAT---SPTTQPAPAT--SPTTQPAPA--TSPTTQPAPATSPTTQPAPATSPTTQPAP 329

Qy    225 ATTPGV-PAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCP 283
Db    330 ATSPTTQPAPATSPTTQP---------------------------------------- 348

Qy    284 EGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLP 343
Db    348 -------------TPA----------TSPTT---------------------------- 356

Qy    344 HGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACT 403
Db    356 ------------------------------------------------------------ 356

Qy    404 GWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPAT 463
Db    356 -----------QPTPATSPT-------------------------TQPTPAT---SPTT 375

Qy    464 QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPA- 523
Db    376 Q--PAPATSP-----------TTQPAPATSPTTQP--APATSPTTQPAPATSPTTQPAP 419

Qy    523 ----TAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPI 578
Db    420 ATSPTTQPAPATSPTTQPAPATSPTTQPTPGT--SPTTQPAPAT-SPTTQPSPA------ 471

Qy    579 TVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAP 638
Db    472 ---TNPTTQPSVSSSASTQPAPATS----GTTQPAPATSATTQPAPATSATTQPAPATSP 523

Qy    639 TTAAPAP-------------NTTVTAPPTAAPTTAAPA---PNTTVTVPPTAAPTAAPP 681
Db    524 TT-QPAPATSATTQSVPVVTTQTTMGTSPTTAGISSSPAATSPSTAVTSPSSLGTSSSP- 582

Qy    682 TVAHAPNTTAAPVTTTS-------APATTP 704
Db    582 -LPSSISTSALPIASSSASSSPSAASSTTP 610


RESULT  11
O76810
ID   O76810      PRELIMINARY;      PRT;   373 AA.
AC   O76810;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ICHIT PROTEIN.
GN   ICHIT.
OS   ANOPHELES GAMBIAE (AFRICAN MALARIA MOSQUITO).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; NEMATOCERA; CULICOIDEA; CULICIDAE; ANOPHELES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SUAKOKO; TISSUE=MAINLY MIDGUT;
RA   DIMOPOULOS G., SEELEY D., WOLF A., KAFATOS F.C.;
RT   "Malaria infection of the mosquito Anopheles gambiae activates immune
RT   responsive genes during critical transition stages of the parasite
RT   life cycle.";
RL   SUBMITTED (SEP-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ010903; E1320052; -.
SQ   SEQUENCE   373 AA;  40647 MW;  A7231CDB CRC32;

  Query Match             9.6%;  Score 422.5;  DB 3;  Length 373;
  Best Local Similarity   35.6%;  Pred. No. 1.7e-06;
  Matches  135;  Conservative   23;  Mismatches  126;  Indels   95;  Gaps   20;

Qy      1 MIKTLLFLTALGL----VAARPEVSDAEKNPALHEPHPDXPPAEQXXLLPXEY----DCT 52
Db      1 MIAAAMKYVALGLVLLAVSARAEPGEVIPN------HPNCP--EMQGPLPHYFIHPTNCS 52

Qy     53 KFYYC------EYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGC--TLPGPPAETTQAPA 104
Db     53 RFYECHMRDAWEY--------ECPAGLHFNVAIDVCDFPVNAKCESQSPGDQTTTTLRPT 104

Qy    105 TTQ-APTTTQA----PTTTTQAPTTTTQAPTTTTQAPTTTQA---PTTTQA------PTT 150
Db    105 TTTLRPTTTTTTDWITTTTTEATTTTKFPTTTTTTSAPTTPSQWTDPTITTTTPVWTDPTT 164

Qy    151 TQAPTTTT------QAPTTTT------QAPTTTTQAPTTTQA--------PTTT------ 185
Db    165 WSAPTTTTTWSDQPPPPTTTTTTVWTDPTATTTTHAPTTTTTWSDLPPPPPTTTTTVWID 224

Qy    185 -QAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATT--------PAATT--PGVPAP 233
Db    225 PTATTTTHAPTTTTTWSDLPPPPPTTTTTTVWTDPTTTTTTDYTTAYPPTTSEPPSTPHP 284

Qy    234 TSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQ 293
Db    285 TD-PHCPPTGATLPN----------YWAHGTDCSRYYGCLEGCVKEFKCPDGLYWNDQQK 333

Qy    294 RCDSPANVECD-GEISPAP 311
Db    334 RCDSYSSSQCGCPDIPPAP 352


RESULT  12
Q41805
ID   Q41805      PRELIMINARY;      PRT;  1188 AA.
AC   Q41805;
```

```
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   EXTENSIN-LIKE PROTEIN PRECURSOR.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;
OC   TRACHEOPHYTA; EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA;
OC   POALES; POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B73; TISSUE=POLLEN;
RA   RUBINSTEIN A.L., BROADWATER A.H., LOWREY K., BEDINGER P.A.;
RL   SUBMITTED (JUN-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z34465; G600118; -.
DR   PFAM; PF00560; LRR; 3.
KW   SIGNAL.
FT   SIGNAL        1     27       POTENTIAL.
SQ   SEQUENCE   1188 AA;  120980 MW;  47E3D9BA CRC32;


  Query Match             9.6%;  Score 421.5;  DB 8;  Length 1188;
  Best Local Similarity   22.7%;  Pred. No. 4e-06;
  Matches  175;  Conservative   89;  Mismatches  323;  Indels  183;  Gaps   29;

Qy       2 IKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDX-----PPAEQXXLLPXEYDCTKFYY 56
           :||      :|  :  | ||     | :  | |       || |:    |
Db     520 VKTTSPPAPIGSPSPPPPVSVVSPPPPVKSPPPPAPVGSPPPPEKSPPPP---------- 570

Qy      57 CEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAP------- 110
                ||    |         | | :       | |||     |   ::|
Db     570 --------APVASPPPPVKSPPPPTLVASPPPPVKSPPPPAPVASPPPPVKSPPPPTPVA 621

Qy     110 -TTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAP 168
            ||  ::  |  :   ||  :  |   ::|       :|  |      ||:  :|
Db     622 SPPPPAPVASSPPPMKSPPPPTPVSSPPPPEKSPPPPPPAKSTPPPEEYPTPPTSVKSSP 681

Qy     169 TTTTQAPTTTQAP------------TTTQAPTTTQAPTTITQAATTPAATTPAATTPA-- 215
           |  |  |           | :: |:: : |:   :  ::|  |  ::  ||
Db     682 PPEKSLPPPTLIPSPPPQEKPTPPSTPSKPPSSPEKPSPPKEPVSSPPQTPKSSPPPAPV 741

Qy     215 -------ATTPAATTPAATTPGV-----PAPTSAPVWPPICELLPNGCPADFDIHLLIPH 262
                  ::| |  | :: | |     ||| |:|  ||  ::      |    :    |
Db     742 SSPPPTPVSSPPALAPVSSPPSVKSSPPPAPLSSP--PPAPQV--KSSPPPVQVSSPPPA 797

Qy     263 DKYCNLFYQCSNGYTFEQRC-PEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDED 321
            |        |:   |:   |     :| :    || :|     :|  ||| : :
Db     798 PKSSPPLAPVSSPPQVEKTSPPPAPLSSPPLAPKSSPPHV----VVSSPPPVVKSSPP-- 852

Qy     322 IDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIE 381
                  ||       | |                :       | |          |
Db     852 -----------PAPVSSPPLTP-----------------KPASPPAHVS------SPPE 876

Qy     382 LVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYL 441
           :|  : |   :  :             |:||    |      |   |:   :   |
Db     877 VVKPSTPPAPTTVISP----------PSEPKSSPP-----PTPVSLPPPIVKSS------ 916

Qy     442 QCVHGQTIARPCPGNLHFSPAT-QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPT 500
                 | |  :   | | :|   ||          |    |  ::|  || ::|
Db     916 ----------PPPAMVSSPPMTPKSSPPPVVV--------SSPPPTVKSSPPPAPVSSPP 957

Qy     501 AAPTAAPTAAPSTVVPP---ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTV 557
           | | ::|  ||  : ||    ::|| |    ||      ||| ::|  || :: |  |  |
Db     958 ATPKSSPPPAPVNLPPPEVKSSPPPTPVSSPP------PAPKSSPPPAPMSSPP--PPEV 1009

Qy     558 TVPPTAAPTAAPTTAV----PEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTS 613
             ||  || ::|   |    |  |::    | :|   |  || ::|       |  | |
Db    1010 KSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSP---PPPAPVSSP-------PPPVKS 1059

Qy     614 PPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPT 673
           ||  || ::  | |   |  ||  || :: | |   | :||  || :: | |   :  ||
Db    1060 PPPPAPISSPPPP---VKSPPPPAPVSSPPPP---VKSPPPPAPVSSPPPP---IKSPPP 1110

Qy     674 AAPTAAPPTVAHAPNTTAAPVTTTS-APATTP-----EDDDIDPPLPNDP 717
```

```
           || ::||     | :   |   :|  |  ||     |:  : ||  : |
Db    1111 PAPVSSPPPAPVKPPSLPPPAPVSSPPPVVTPAPPKKEEQSLPPPAESQP 1160


RESULT  13
Q17921
ID   Q17921      PRELIMINARY;      PRT;   825 AA.
AC   Q17921;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   C12D12.1 PROTEIN.
GN   C12D12.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA   BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA   DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA   HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA   KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA   MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA   RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA   SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA   VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA   WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   NHAN M., HAWKINS J.;
RL   SUBMITTED (APR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U51998; G1255319; -.
SQ   SEQUENCE   825 AA;  86617 MW;  90FA8FCE CRC32;


  Query Match             9.4%;  Score 414.5;  DB 3;  Length 825;
  Best Local Similarity   23.8%;  Pred. No. 5e-06;
  Matches  150;  Conservative   27;  Mismatches  162;  Indels  291;  Gaps   17;

Qy     102 APATTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAP 161
           || ||  ||     | :| :| ||   ||| |  |||| |    |  ||  |||   :
Db     460 APPTTPGPTFYPVTTVSTMSPPTTVTVPTTPTPVPTTTNTP---PANPTTATPTTVGTSK 516

Qy     162 TTTTQAP-----------TTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAA 210
               | | :|           :| | || |: :||||   ||   |||     | |  || :
Db     517 QTNTISPHLSTITGSIVTSTPTMAPQTSASPTTTPTHTTASQPTT-----TKPVVTTNSV 571

Qy     211 TTPAATTPAATTPAATTPGVP-APTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLF 269
           |    ||   | |  || | |   |:|||  |
Db     572 TPSTGTT---TVPVPTTTGSPTTQTTAPVTKP--------------------------- 601

Qy     270 YQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLD 329

Db     601 ------------------------------------------------------------ 601

Qy     330 NGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPG 389
                                                                  |:|
Db     601 ----------------------------------------------------TVP- 604

Qy     390 GESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTI 449
```

```
Db    604 ------------------------------------------------------------ 604

Qy    450 ARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTA 509
                                                 |:  || |  ||: |    |:
Db    604 ------------------------------------SSTTQTAPPVTTPTSQPPVTTTS 626

Qy    510 APSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTA-APESPTTVTVPPTAAPTAA 568
          :|:  | | | |   || :  :|| ||    || ||: ||     ||  || ||| :
Db    627 LLTTLTTP-TVPVTTTVVPSSATVPTTPPTTVTVAATTTSKAP----VVTTSPTLAPT-S 680

Qy    569 PTTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAA--PTTAAPAP 626
          ||      | || ::|||       |    |||| |. ||: |: |||   ||    :|
Db    681 PTKLPTSPPSTVGTSPTA-------PANLTTPTTA-PVNPTSSTTAPTAPVNPT----SP 728

Qy    627 NTTVTVPP--TAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVA 684
          |  ||||  |  |||      || |    |    |    |    |  |: || |   |   |
Db    729 TTAPTVPPVTTTTPTTTTSTTTTTTTTTTTTQTTPTTPVVTTPSTITPTTRPVTQP---A 785

Qy    685 HAPNTTAAPVTTTSAPATTPEDDDIDPPLP 714
          | |  |   ||: | ||     :   ||
Db    786 SQPATQPATTPTTTIPTTTQFSSAVSVALP 815


RESULT  14
O42854
ID   O42854      PRELIMINARY;      PRT;  1611 AA.
AC   O42854;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 170.5 KD PROTEIN.
GN   SPAC23A1.16.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   MURPHY L., HARRIS D.;
RL   SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   WOOD V.; BARRELL B.G., RAJANDREAM M.A.;
RL   SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AL021813; E1250590; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   1611 AA;  170526 MW;  98811CA3 CRC32;

  Query Match             9.4%;  Score 412.5;  DB 1;  Length 1611;
  Best Local Similarity   22.3%;  Pred. No. 8.8e-06;
  Matches  186;  Conservative   80;  Mismatches  324;  Indels  243;  Gaps   37;

      19 EVSDAEKNPALHE---PHPDXPPAEQXXLLPXEYDCTKFYYCEYGLKFIAPRDCAPGTEF 75
         | |:|:: |:  |   | |  |   |    | ||  |            ||:  |
     492 EQSNAQQAPSPKEEERPLPSEPSQNQ----PAEYRDTP----------DTPRNIMP---- 534

      76 KFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQA 135
                 | |       | : |:          : |    :     |: |
     534 ----------LPGLMSADQPIKVTEPSNDADKAIVAEGPNNEEE-----TKGPVIPETQ 577

     136 PTTTQAPTTTQAPTTTQA---PTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQA 192
          |: |    | :|   :    |   |    . |  :        |    ||  |:      :
     578 ETSEQQVHKTPSPEKQKVLSPPPIITNFDKETLASNEAHEAVPQKPSAPQVTRLMAPQDS 637

Qy   193 PTTITQAATT---PAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNG 249
          : :| : |:   ||             |     |  |  | :  ::|: ||     |:
Db   638 SSVVTPSPTSLLDPARAVRKVID--GIDPPKEAGAGATADVESAANSPITPPRTWHSPDF 695

Qy   250 CPADFD-IHLLIP------------HDKYCNLFYQCSNGYTFEQRCPEGLYFN------- 290
              |: |   :|           ||      :     : | || |
```

```
Db    696 TSKSFEPIERKLPSRISEVTEDSIDEDKQ----NEVDPSTSARALPPPGLRFGKVDTLAS 751

Qy    290 ---------PYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDW 340
                  | | |  ||          |  | |          |:  ||        ::
Db    752 LAHDDLDDLPAVPRIFSP----------PPLPKTPS--------GEFGDN--------EF 785

Qy    341 LLP-HGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDE 399
          : |   ||        | |:  : |   |:       |  ::  |     ||
Db    786 MFPKKSNR-------VRGH--QSRPSTGS-------QLRNVVPVSIVTSGGR-------- 822

Qy    400 DACTGWYCPTEPIEW--------EPLPNGCPADF---SIDHLLPHESDCGQYL------- 442
                | | |           |||:  ||||   ::|   ||      ||
Db    822 --------PALPDEMASPSSSIGHPLPSPPPADFNSLNVDFYEPH-----SYLESPAPEP 868

Qy    442 ------QCVHGQTIARPCPGNLHF--------------SPATQSCESPVTAGCQVFECD 480
                :  :   |  | |  |                    |:|  |||       :
Db    869 QPSYEEESFNATVIHAPTPSTATFQGHPTISNVATPPLKQDVTESKASPVADASATHQSS 928

Qy    481 S--DNQCTSTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVP-------PTT 531
          :   : |   :   ||    :      : |   ||: ||    |:|     ||:
Db    929 TGLTQEITQLGSNMRLPTKLTRPSNDGRKASGPRPAAPPSIPP----PLPVSNILSSPTS 984

Qy    532 AIPTPAPTAAPTAAPTTAAPESPTTVTVPPT------AAPTAAPTTAVPEIPITVTSAPT 585
            |  | :|| : | : :| :|    |||       | |   |:   | ||:  |: |
Db    985 EPPKDHPPSAPLSKPVSTSPAAP-LARVPPVPKLSSKAPPVPLPSADAPPIPVPSTAPPV 1043

Qy    586 AAPTAAP------TAAPTA-----APTTAVPEIPTTVTSPPTAAPT----------TAAP 624
            ||: |      : ||:|     ||:: :| ||    :||  ||:          | |
Db   1044 PIPTSTPPVPKSSSGAPSAPPPVPAPSSEIPSIPAPSGAPPVPAPSGIPPVPKPSVAAPP 1103

Qy    625 APNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVA 684
            |  :| |||  ||: | | |  :| |||    |: | | |  :|  ||  ||: |||
Db   1104 VPKPSVAVPPVPAPSGAPPVPKPSVAAPPVPVPSGAPPVPKPSVAAPPVPAPSGAPP--V 1161

Qy    685 HAPNTTAAPVTTTSA------------PATTPEDDDIDPPLPN-----DPINP 720
            |:  | ||   |:            |   |  :   ||:|       |: |
Db   1162 PKPSVAAPPVPAPSSGIPPVPKPAAGVPPVPPPSE--APPVPKPSVGVPPVPP 1212


RESULT  15
O13028
ID   O13028      PRELIMINARY;      PRT;   507 AA.
AC   O13028;
DT   01-JUL-1997 (TREMBLREL. 04, CREATED)
DT   01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ANTIFREEZE GLYCOPEPTIDE AFGP POLYPROTEIN PRECURSOR.
OS   BOREOGADUS SAIDA.
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; PARACANTHOPTERYGII; GADIFORMES; GADOIDEI;
OC   GADIDAE; BOREOGADUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97268653.
RA   CHEN L., DEVRIES A.L., CHENG C.H.C.;
RT   "Convergent evolution of antifreeze glycoproteins in Antarctic
RT   notothenioid fish and Arctic cod.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 94:3817-3822(1997).
DR   EMBL; U43200; G2078483; -.
KW   SIGNAL; POLYPROTEIN.
FT   SIGNAL        1     33       POTENTIAL.
FT   CHAIN        34    507       ANTIFREEZE GLYCOPEPTIDE AFGP POLYPROTEIN.
SQ   SEQUENCE   507 AA;  44684 MW;  675D6FCB CRC32;


  Query Match             9.2%;  Score 405;  DB 12;  Length 507;
  Best Local Similarity   25.5%;  Pred. No. 6.8e-06;
  Matches  172;  Conservative   17;  Mismatches  256;  Indels  230;  Gaps   22;

Qy     87 LAGCTLPGP----------------PAETTQA--PATTQAPTTTQAPTTTTQAPTTTTQ 127
          | |  | ||                ||   :|  |||   | |   | |   | |  |:
Db      2 LTGSCLLGPSCTSSPRYEQHQLLVARPAAAARAATPATAATPATAATPAT---AATAATE 58
```

```
Db    128 APTTTTQAPTTTQAPTTTQAPTTTQA----------PTTTTQAPTTTTQA-PTTTTQAPT 176
          |   |   | |   | |     || |           |    :| |  | | | |   | |
      59  ATAATAATPATAATPATAATAATTAATAATAATAATPARAARAATPATAATPATAATAAT 118

      177 -----TTQAPTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPA-ATTPGV 230
               | : |     | |   | |    ||   | | | :  |||   | ||| | ||
          1   ATAATAETPARAATPATAATPATAATPATAATAATAATSATAATAARAATPATAATPAT 178

          PAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNP 290
          ||  :
      79  PATAAR---------------------------------------------------- 185

      291 YVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDK 350
             |  :||        : |   |                 ||
      185 -AARAATPA--------TAATAATAATAATAATAATAATAATPA---------------- 220

      351 YYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTE 410
                     |   | |  :                                    ||
Db    220 -----------RAARAATPAT-----------------------------------APT- 233

Qy    411 PIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPV 470
                |      ||  :     |              | | |       :||| :
Db    233 -----PATAATPATAATAATAP--------------TAATPARAARAATPATAA----- 268

Qy    471 TAGCQVFECDSDNQCTSTAAPTAAPTAAPTAA--PTAAPTAAPSTVVPPATPPATAAP-V 527
                          :|||  | |    |||   |||  | |:    ||||   | |
Db    268 --------------TLATAATPATPATPATAATDATAATAATPARAATPATPATAATPAT 313

Qy    528 PPTTAIPTPAPTAAPTAAPTTAA-PESPTTVTVPPTAAPTA-------APTTAVP----- 575
          | | |    | |||  | |  ||   :| |   | |||  |       | | | |
Db    314 PATAATAATAATAATAATPARAARAATPATAATPATAATAATAATAATAATAATPARAAR 373

Qy    575 -EIPITVTSAPTA--APTAAPTAAPTAAPTTAVPEIPTTVTSP--PTAAPTTAAPAPNTT 629
            | |  :| ||  | |||  | |  |   | |  | |  :|  |  | | |  |   |
Db    374 AATPATAATAATAATAATAATAATPARAARAATPATPATPATPATPATAATAATAATAAT 433

Qy    630 VTVPPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNT 689
             |||    || |     |   |   | | ||          | ||| |  |  | |
Db    434 AATAATAATAATAPTPARAARAATPATGATPATAPTAGTAATAATAATAATPARASTPAT 493

Qy    690 TAAPVTTTSAPATTP 704
           | |||     | ||
Db    494 AATPVT-----AATP 503
```

Search completed: July 18, 1999, 02:42:07
Job time: 3630 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 01:41:37 ; Search time 32.5 Seconds
(without alignments)
1337.641 Million cell updates/sec

Title: US-09-294-663-3
Perfect score: 4328
Sequence: 1 MIKTLLFLTALGLVAARPEV.....VFIPWNKLDEDIRQALNFEL 788

Scoring table: BLOSUM62

Searched: 180763 seqs, 55169189 residues

Database : SPTREMBL_8:*
1: sp_fungi:*
2: sp_human:*
3: sp_invertebrate:*
4: sp_mammal:*
5: sp_mhc:*
6: sp_organelle:*
7: sp_phage:*
8: sp_plant:*
9: sp_bacteria:*
10: sp_rodent:*
11: sp_virus:*
12: sp_vertebrate:*
13: sp_unclassified:*
14: sp_archea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4328	100.0	788	3	O18510	O18510 trichoplusi
2	4298.5	99.3	807	3	O18511	O18511 trichoplusi
3	520.5	12.0	1404	2	Q92954	Q92954 homo sapien
4	461.5	10.7	3570	2	Q99552	Q99552 homo sapien
5	461.5	10.7	851	3	O17893	O17893 caenorhabdi
6	457	10.6	867	11	O39782	O39782 equine herp
7	452.5	10.5	866	11	O39781	O39781 equine herp
8	448.5	10.4	770	3	Q20908	Q20908 caenorhabdi
9	447.5	10.3	801	3	Q23635	Q23635 caenorhabdi
10	442.5	10.2	373	3	O76810	O76810 anopheles g
11	414.5	9.6	2378	3	P91365	P91365 caenorhabdi
12	408	9.4	825	3	Q17921	Q17921 caenorhabdi
13	393	9.1	1777	3	Q22579	Q22579 caenorhabdi
14	392	9.1	1188	8	Q41805	Q41805 zea mays (m
15	388	9.0	750	11	O39307	O39307 equine herp
16	387.5	9.0	622	2	Q14881	Q14881 homo sapien
17	383	8.8	507	12	O13028	O13028 boreogadus
18	371	8.6	1611	1	O42854	O42854 schizosacch
19	367	8.5	1151	12	O57580	O57580 gallus gall
20	366	8.5	1131	3	Q18529	Q18529 caenorhabdi
21	362	8.4	412	3	O77242	O77242 heterodera
22	361.5	8.4	1118	4	Q28226	Q28226 canis famil
23	351.5	8.1	2187	10	P70670	P70670 mus musculu
24	351	8.1	846	3	O01699	O01699 caenorhabdi
25	345	8.0	232	3	Q27423	Q27423 drosophila
26	340.5	7.9	353	3	Q25677	Q25677 podocoryne
27	340.5	7.9	584	3	Q17802	Q17802 caenorhabdi
28	340	7.9	279	2	Q14888	Q14888 homo sapien
29	339.5	7.8	901	9	Q44562	Q44562 actinomyces
30	328	7.6	379	3	Q27929	Q27929 drosophila
31	327	7.6	2225	3	O45881	O45881 caenorhabdi
32	327	7.6	2761	3	Q19522	Q19522 caenorhabdi
33	326	7.5	5376	10	O88799	O88799 mus musculu
34	324	7.5	1229	3	Q94185	Q94185 caenorhabdi
35	324	7.5	528	4	Q29071	Q29071 sus scrofa
36	323	7.5	928	12	Q98906	Q98906 gallus gall
37	323	7.5	722	12	O13083	O13083 dissostichu
38	322	7.4	216	4	Q28501	Q28501 macaca mula
39	321.5	7.4	350	8	Q41719	Q41719 zea diplope
40	314.5	7.3	303	8	Q41814	Q41814 zea mays (m
41	314.5	7.3	3247	11	Q65553	Q65553 bovine herp
42	314	7.3	328	8	Q42366	Q42366 zea mays (m
43	313	7.2	946	8	O22015	O22015 cylindrothe
44	313	7.2	907	11	Q66537	Q66537 epstein-bar
45	310	7.2	886	8	O22016	O22016 cylindrothe

ALIGNMENTS

RESULT 1
O18510
ID O18510 PRELIMINARY; PRT; 788 AA.
AC O18510;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE INSECT INTESTINAL MUCIN IIM14.
OS TRICHOPLUSIA NI (CABBAGE LOOPER).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; LEPIDOPTERA; NOCTUOIDEA; NOCTUIDAE; TRICHOPLUSIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97341213.
RA WANG P., GRANADOS R.R.;
RT "Molecular cloning and sequencing of a novel invertebrate intestinal
RT mucin cDNA.";
RL J. BIOL. CHEM. 272:16663-16669(1997).
DR EMBL; AF000605; G2224919; -.
SQ SEQUENCE 788 AA; 81716 MW; 8F8ADE83 CRC32;

Query Match 100.0%; Score 4328; DB 3; Length 788;
Best Local Similarity 100.0%; Pred. No. 4.2e-120;
Matches 788; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60

Qy      61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy     121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQA 180

Qy     181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240

Qy     241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300

Qy     301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360

Qy     361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420

Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAA 660

Qy    661 PPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHADCDKYW 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 PPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINPCVEECNVLPWAHADCDKYW 720

Qy    721 VCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMSESYEGVQVFIPWNKLDEDI 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 VCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMSESYEGVQVFIPWNKLDEDI 780

Qy    781 RQALNFEL 788
          ||||||||
Db    781 RQALNFEL 788
```

RESULT 2
O18511
ID O18511 PRELIMINARY; PRT; 807 AA.
AC O18511;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE INSECT INTESTINAL MUCIN IIM22.
OS TRICHOPLUSIA NI (CABBAGE LOOPER).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; LEPIDOPTERA; NOCTUOIDEA; NOCTUIDAE; TRICHOPLUSIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97341213.
RA WANG P., GRANADOS R.R.;
RT "Molecular cloning and sequencing of a novel invertebrate intestinal
RT mucin cDNA.";
RL J. BIOL. CHEM. 272:16663-16669(1997).
DR EMBL; AF000606; G2224921; -.
SQ SEQUENCE 807 AA; 83408 MW; 15E08C34 CRC32;

Query Match 99.3%; Score 4298.5; DB 3; Length 807;
Best Local Similarity 97.5%; Pred. No. 3.1e-119;
Matches 787; Conservative 0; Mismatches 1; Indels 19; Gaps 1;

```
Qy      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MIKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYG 60

Qy     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQ 120

Qy    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 APTTTTQAPTTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQA 180

Qy    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 PTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWP 240
```

```
Qy    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 PICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPAN 300

Qy    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360

Qy    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNG 420

Qy    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECD 480

Qy    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTA 540

Qy    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 APTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAPTAAP 600

Qy    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTV------------------PPTAAPTTA 641
          ||||||||||||||||||||||||||||||||                  |||||||||
Db    601 TTAVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTAPPTAAPTTA 660

Qy    642 APAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 701
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    661 APAPNTTVTVPPTAAPTAAPPTVAPAPNTTAAPVTTTSAPATTPEDDDIDPPLPNDPINP 720

Qy    702 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CVEECNVLPWAHADCDKYWVCDGNNQVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 780

Qy    762 ESYEGVQVFIPWNKLDEDIRQALNFEL 788
          ||||||||||||||||||||||||||||
Db    781 ESYEGVQVFIPWNKLDEDIRQALNFEL 807
```

RESULT 3
Q92954
ID Q92954 PRELIMINARY; PRT; 1404 AA.
AC Q92954;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MEGAKARYOCYTE STIMULATING FACTOR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA TURNER K.J., FITZ L.J., TEMPLE P., JACOBS K., LARSON D., LEARY A.C.,
RA KELLEHER K., GIANNOTTI J., CALVETTI J., FITZGERALD M., KRIZ M.J.,
RA FERENZ C., GROBHOLZ J., FRASER H., BEAN K., NORTON C.R., GESNER T.,
RA BHATIA S., KRIZ R., HEWICK R., CLARK S.C.;
RL BLOOD 78:279-279(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA MERBERG D.M., FITZ L.J., TEMPLE P., GIANNOTTI J., MURTHA P.,
RA FITZGERALD M., SCALTRETO J., KELLEHER K., PREISSNER K., KRIZ R.,
RA JACOBS K., TURNER K.;
RL (IN) PREISSNER K.T., ROSENBLATT S., KOST C., WEGERHOFF J.,
RL MOSHER D.F. (EDS.); BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS.:45-52;
RL ELSEVIER SCIENCE PUBLISHERS B.V. (1993).
RN [3]
RP SEQUENCE FROM N.A.
RA TURNER K.J., FITZ L.J., TEMPLE P., JACOBS K., LARSON D., LEARY A.C.,
RA KELLEHER K., GIANNOTTI J., CALVETTI J., FITZGERALD M., KRIZ M.J.,
RA FERENZ C., GROBHOLZ J., FRASER H., BEAN K., NORTON C.R., GESNER T.,
RA BHATIA S., KRIZ R., HEWICK R., CLARK S.C.;

```
RL   SUBMITTED (SEP-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U70136; G1572721; -.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR   PFAM; PF00045; hemopexin; 2.
DR   PFAM; PF01033; Somatomedin_B; 2.
SQ   SEQUENCE   1404 AA;  151090 MW;  6C1F4DEA CRC32;


  Query Match             12.0%;  Score 520.5;  DB 2;  Length 1404;
  Best Local Similarity   29.5%;  Pred. No. 4.1e-09;
  Matches  203;  Conservative   26;  Mismatches  205;  Indels  253;  Gaps   32;

Qy      93 PGPPAE-TTQAPATT----QAPTTTQAPTTTT----------QAPTTTTQ-APTTTTQAP 136
           | | || ||: || |      ||| : | :||          ||||  : |||||  ||
Db     331 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAP 390

Qy     137 TTTQ--APTTTQ--APTTTQ--APTTTTQ-APTTTTQAPTTTTQ-APTTTQ--APTTTQ- 186
           || :  |||||:  |||| :  ||||| : |||||  ||||  : |||| :  |||| :
Db     391 TTPKEPAPTTTKEPAPTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKKPAPTTPKE 450

Qy     186 -APTTTQAPTTITQAATTPAATTPAATTPAATTPAA-TTPAATTPGVPAPTSAPVWPPIC 243
            |||| : ||  |      |    || |||    | |   || |||  ||||: |  |
Db     451 PAPTTPKEPTPTTPKEPAPTTKEPAPTTPKEPAPTAPKKPAPTTPKEPAPTT-PKEPA-- 508

Qy     244 ELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVEC 303
                                                  |   :  ||   :
Db     508 ----------------------------------------------PTTTKEPSPTTPK- 521

Qy     304 DGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
               |||  |
Db     521 ----EPAPTTT-------------------------------------------------- 528

Qy     364 CGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDA-CTGWYCPTEPIEWEPLPNGCP 422
                                  :|      | |  |    || | |  |     |
Db     528 --------------------------KSAPTTTKEPAPTTTKSAPTTPKEPSPTTTKEP 560

Qy     423 ADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSD 482
           |  :    | |                          :| |    :| |
Db     561 APTT-----PKEP-----------------------APTTPKKPAPTT----------- 581

Qy     483 NQCTSTAAPTAAPTAAPTAAPTAAPTA--APSTVVPPATPPATAAPVPPTT---AIPT-- 536
                |||     |||      ||||   |:   |  | | |    : |||      ||
Db     581 ---PKEPAPTTPKEPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTP 637

Qy     536 --PAPTAAPTAAPTTAAPESPTTVTVP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589
             ||||     ||||     :|||    | ||    ||| |  |: |    |    |||
Db     638 EKPAPTTPEELAPTTPEEPTPTTPEEPAPTTPKAAAPNT--PKEPAPTTPKEPAPTTPKE 695

Qy     590 AAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP---APNTT------------VTVPP 634
             |||      ||||      |||:  |     |    ||   || ||             | |
Db     696 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPK 755

Qy     635 TAAPTT-AAPAPNTTVTVPPTAAPTAAPPTVAH-------------APNTTAAPVTTTS 679
             ||||   ||| |      ||      || |:                || ||  | :|||
Db     756 GTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTS 815

Qy     680 ---APATTPEDDDIDP--PLPNDPINP 701
              || |  |      |  | |  |  |
Db     816 DKPAPTTPKETAPTTPKEPAPTTPKKP 842


RESULT   4
Q99552
ID   Q99552      PRELIMINARY;      PRT;  3570 AA.
AC   Q99552;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MUCIN MUC5B (FRAGMENT).
GN   MUC5B.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.

RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 97166151.
RA   DESSEYN J.L., GUYONNET-DUPERAT V., PORCHET N., AUBERT J.P., LAINE A.;
RT   "Human mucin gene MUC5B, the 10.7-kb large central exon encodes
RT   various alternate subdomains resulting in a super-repeat. Structural
RT   evidence for a 11p15.5 gene family.";
RL   J. BIOL. CHEM. 272:3168-3178(1997).
DR   EMBL; Z72496; E244318; -.
FT   NON_TER       1      1
FT   NON_TER    3570   3570
SQ   SEQUENCE   3570 AA;  361016 MW;  452C733D CRC32;


  Query Match             10.7%;  Score 461.5;  DB 2;  Length 3570;
  Best Local Similarity   24.8%;  Pred. No. 4e-07;
  Matches  203;  Conservative   82;  Mismatches  254;  Indels  279;  Gaps   39;

Qy      72 GTEFKFSAQTCVHAALAGCTLPGPPAET--TQAPATTQA----------PTTTQAPTTTT 119
           || :   :||    |       |    ||  |    || |          |::|     ||
Db    1892 GTAWTRLSQTTTPTATMSTATPSSTPETVHTSTVLTTTATTTRTGSVATPSSTPGTAHTT 1951

Qy     120 QAPTTTTQAPTTT-TQAPTTTQAP----TTTQAPTT---TQAP--------------TT 156
           : |||||   | | : :| |   |    :||  |||   |  |                ||
Db    1952 KVPTTTTTGFTATPSSSPGTALTPPVWISTTTTPTTRGSTVTPSSIPGTTHTATVLTTTT 2011

Qy     157 TTQA------PTTTTQA----PTTTTQAPTTTQAPTTT---QAPTTTQAPTTITQAATTP 203
           || |      |:::||     |: || | | |   :||      | ||  |  :|  ||||
Db    2012 TTVATGSMATPSSSTQTSGTPPSLTTTATTITATGSTTNPSSTPGTTPIPPVLTTTATTP 2071

Qy     204 AATT--------------PAATTPAAT------------------------------ 217
           |||:                |   | | |
Db    2072 AATSSTVTPSSALGTTHTPPVPNTTATTHGRSLPPSSPHTVPTAWTSATSGILGTTHITE 2131

Qy     217 -------TPAATT-------------------------PAATTPG------------VPA 232
                   ||||||                        |  ||||            |:
Db    2132 PSTGTSHTPAATTGTTQPSTPALSSPHPSSRTTESPPSPGTTTPGHTRGTSRTTATATPS 2191

Qy     233 PT-------SAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCS----NGYTFEQR 281
           |       |:|   ||  ::  ||                        ||:      |::
Db    2192 KTRTSTLLPSSPTSAPITTVVTTGCEP-----------------QCAWSEWLDYSYPMP 2233

Qy     282 CPEGLYFNPYVQ-------RCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPA 334
           | |  |: |          |: |  :||  :  |  |: |         :| :::  |
Db    2234 GPSGGDFDTYSNIRAAGGAVCEQPLGLECRAQAQPGVPLRE--------LGQVVE--CSL 2283

Qy     335 NFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEE 394
           :| :  :  :  :  |:  |          |::|::       | |
Db    2284 DFGL--VCRNREQVGKFKMC--------------FNYEIR-------VFC---------- 2311

Qy     395 VDVDEDACTGWYCPTEP-IEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPC 453
                 |   :||: |         |:  |     |    |   :    :       | :
Db    2311 -------CNYGHCPSTPATSSTATPSSTPGTTWILTKLTTTATTTE-------STGSTAT 2356

Qy     454 PGNLHFSPA-----TQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTA-APTAAP-TAA 506
           | :    ||      : :  :|      :      |    |:|| |    ||  |||   ||
Db    2357 PSSTQGPPAGTPHVSTTATTPTVTSSKATPFSSPG--TATALPALRSTATTPTATSFTAI 2414

Qy     507 P-------------TAAPSTVVPPATPPATAAPVPPTTAIPTPAPT-------AAPTAAP 546
           |                ' |   :  ||| :|    |  :| : | | |        | |:: |
Db    2415 PSSSLGTTWTRLSQTTTPMATMSTATPSSTPETVHTSTVLTTTATTTGATGSVATPSSTP 2474

Qy     547 TTA-APESPTTVTVPPTAAPTAAPTTA-VPEIPITVTSAPTAAPTAAPTAAPTAAP-TTA 603
           ||   : ||| |    |  |:::| ||  | : |: |: ||   |:  |  |:: | ||
Db    2475 GTAHTTKVPTTTTTGFTVTPSSSPGTARTPPVWISTTTTPT---TSGSTVTPSSIPGTTH 2531

Qy     604 VPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAP--TAAP 661
           | : || |: | |  : | |: :|  :  | :  |||      |  |  |:: |  |  |
Db    2532 TPTVLTT-TTQPVATGSMATPSSSTQTSGTPPSLITTATTITATGSTTNPSSTPGTTPIP 2590

Qy     662 P---TVAHAPNTTAAPVTTTSAPATTPEDDDIDPPLPN 696
           |   | |  |  |:: || :||  ||       ||:||
```

```
Db   2591 PELTTTATTPAATSSTVTPSSALGTTH-----TPPVPN 2623


RESULT   5
O17893
ID   O17893      PRELIMINARY;      PRT;   851 AA.
AC   O17893;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   F55B11.3.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA   BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA   DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA   HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA   KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA   MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA   RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA   SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA   VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA   WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
DR   EMBL; Z83318; E1186497; -.
SQ   SEQUENCE   851 AA;  92127 MW;  87CF937F CRC32;


  Query Match             10.7%;  Score 461.5;  DB 3;  Length 851;
  Best Local Similarity   25.3%;  Pred. No. 1.5e-07;
  Matches  151;  Conservative   17;  Mismatches  137;  Indels  291;  Gaps   15;

Qy      91 TLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTT 150
           | |  | : |  |:   : ||   ||||| ||||| |||||  |||    ||| : ||
Db     320 TRPTIPTKLTTTPSEL-STTTASVPTTTTSVPTTTTTVPTTTTTVPTT----TTTVSTTT 374

Qy     151 TQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAA 210
           |  ||||| |||||   ||||  |||    |||  |||  ||| |         | |
Db     375 TTVPTTTTTVPTTTTTVSTTTTTVPTT----TTTVPTTTTTVPTTTT--------TVPTT 422

Qy     211 TTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFY 270
           ||   ||   ||   ||  ||  |:
Db     423 TTSVPTT--TTTVPTTTTTVPTTTT---------------------------------- 446

Qy     271 QCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDN 330

Db     446 ------------------------------------------------------------ 446

Qy     331 GCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGG 390

Db     446 ------------------------------------------------------------ 446

Qy     391 ESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIA 450
                                                                      |::
Db     446 ------------------------------------------------------------TVS 448

Qy     451 RPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAA 510
                                                 |:|  ||   | : |       |
Db     449 ---------------------------------TTTTTVPTTTTTVSTTT------TTV 468

Qy     511 PSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPT 570
           |:|     | | |    || ||   |  ||    |: |||      || ||| |     :  |
Db     469 PTTT---TTVPTTTTTVPTTT---TTVPTTT-TSVPTTTTTVPTTTTTVPTTTTVSTTT 521

Qy     571 TAVPEIPITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTV 630
           | |   | | |: ||   | : |       || ||   |||::  |  |||   || ||
Db     522 TTV---PTTTTTVPTTTTTVSTTTTTVPTTTTTVPTTTTTVSTTTTTVPTTTTAPTTTT 578
```

```
Qy     631 TVP--PTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTSAPATT 684
           |||   |  |||      || ||| |   |  | |      ||  | |||: | ||
Db     579 TVPTTTTTVPTTTTTVSTTTTTVPTTT--TTVPTT------TTTVPTTTTTVPTTT 626


RESULT   6
O39782
ID   O39782      PRELIMINARY;      PRT;   867 AA.
AC   O39782;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE   MEMBRANE GLYCOPROTEIN.
OS   EQUINE HERPESVIRUS 1.
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC   ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BK343;
RA   KIRISAWA R., KOBAYASHI T., KAWAKAMI Y., IWAI H.;
RL   J. EQUINE SCI. 7:79-87(1996).
DR   EMBL; D88734; D1020846; -.
KW   MEMBRANE.
SQ   SEQUENCE   867 AA;  86630 MW;  DC3E8AA2 CRC32;


  Query Match             10.6%;  Score 457;  DB 11;  Length 867;
  Best Local Similarity   27.5%;  Pred. No. 2.1e-07;
  Matches  182;  Conservative   31;  Mismatches  150;  Indels  298;  Gaps   20;

Qy      78 SAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTT--TTQA 135
           | |:   || :  : |   : ||  | :|   |||  || :|  ||||| |||| || |
Db      77 STQSSSTAATSS-SAPSTASSTTSIPTSTSTETTTTTPTASTTTPTTTTAAPTTAATTTA 135

Qy     136 PTT---TQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTT----------TTQAPTTTQAPT 182
            ||   | | |||   | |  |||||   |||| | ||         || | ||| | |
Db     136 VTTAASTSAETTTATATATSTPTTTTPTSTTTTTATTTVPTTASTTTDTTTAATTTAATT 195

Qy     183 ---TTQAPTTTQAPTT--ITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAP 237
               || | ||| | ||   | |||| |||| |||| |||| |||| |||    | |:|
Db     196 TAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATT--TAATTTA- 253

Qy     238 VWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDS 297

Db     253 ------------------------------------------------------------ 253

Qy     298 PANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHG 357

Db     253 ------------------------------------------------------------ 253

Qy     358 NLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPL 417

Db     253 ------------------------------------------------------------ 253

Qy     418 PNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVF 477
                                                        || |  |  ||
Db     253 ------------------------------------------ATTS--SATTA----- 262

Qy     478 ECDSDNQCTSTAAPTAAPT---AAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIP 534
                  |:||| | | |    |  ||| | | |   :|     |  ||      | |
Db     262 -------ATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATT 314

Qy     535 TPAPT--AAPTAAPTTAAPESPTTVTVPPTAAPTAAP-----TTAVPEIPITVTSAPTAA 587
           | | |  |  ||| || :| | :| |   | | |: |    |:| |    |  :| |:
Db     315 TAATTTAATTTAATTTGSPTSGSTST---TGASTSTPSASTATSATPTSTSTSAAATTST 371

Qy     588 PTAAPTAAPTAAPTT--AVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTA---- 642
           ||  ||:| |:| :|  |    ||| |: |: | |      :|||:   |:| |||
Db     372 PT--PTSAATSAESTTEAPTSTPTTDTTTPSEATTATTSPESTTVSASTTSATTTAFTTE 429

Qy     642 ---APAPNT--TVTVPPTAAPTAAPPTVAHAPNTTAAPVTT--TSAPATTPEDDDIDPPL 694
              :|  :|  | |  |::  |  |        :|| | |   | : |:|  |      :
```

```
Db    430 SHTSPDSSTGSTSTAEPSSTFTLTP--------STATPSTDQFTGSSASTESDSTDSSTV 481

Qy    695 P 695
          |
Db    482 P 482
```

RESULT 7
O39781
ID O39781 PRELIMINARY; PRT; 866 AA.
AC O39781;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE MEMBRANE GLYCOPROTEIN.
OS EQUINE HERPESVIRUS 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HH1;
RA KIRISAWA R., KOBAYASHI T., KAWAKAMI Y., IWAI H.;
RL J. EQUINE SCI. 7:79-87(1996).
DR EMBL; D88733; D1020845; -.
KW MEMBRANE.
SQ SEQUENCE 866 AA; 86463 MW; 7200333A CRC32;

Query Match 10.5%; Score 452.5; DB 11; Length 866;
Best Local Similarity 26.8%; Pred. No. 2.8e-07;
Matches 174; Conservative 33; Mismatches 166; Indels 277; Gaps 18;

```
Qy     78 SAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTT--TTQA 135
Db     77 STQSSSTAATSS-SAPSTASSTTSIPTSTSTETTTTTPTASTTTPTTTTAAPTTAATTTA 135

Qy    136 PTT-------TQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTQAPTT-TTQAP 187
Db    136 VTTAASTAASTSAETTTATATATSTPTTTTPTSTTTT---TATTTVPTTASTTTDTTTAA 192

Qy    188 TTTQAPTT--ITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICEL 245
Db    193 TTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATT--TAATTTA--------- 242

Qy    246 LPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDG 305
Db    242 ------------------------------------------------------------ 242

Qy    306 EISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCG 365
Db    242 ------------------------------------------------------------ 242

Qy    366 AGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADF 425
Db    242 ------------------------------------------------------------ 242

Qy    426 SIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQC 485
Db    242 ----------------------------------ATTSSATTAATTTAATTTAATTTAA 266

Qy    486 TSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPT--AAPT 543
Db    267 TTTAATTTAATT--TAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 324

Qy    544 AAPTTAAPESPTTVTVPPTAAPTAAP-----TTAVPEIPITVTSAPTAAPTAAPTAAPTA 598
Db    325 AATTTGSPTSGSTST---TGASTSTPSASTATSATPTSTSTSAAATTSTPT--PTSAATS 379

Qy    599 APTT--AVPEIPTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTA-------APAPNT-- 648
Db    380 AESTTEAPTSTPTTDTTTPSEATTATTSPESTTVSASTTSATTTAFTTESHTSPDSSTGS 439

Qy    648 TVTVPPTAAPTAAPPTVAHAPNTTAAPVTT--TSAPATTPEDDDIDPPLP 695
```

```
          | |  |::  |  |        :|| | |   | : |:|  |      :|
Db    440 TSTAEPSSTFTLTP--------STATPSTDQFTGSSASTESDSTDSSTVP 481
```

RESULT 8
Q20908
ID Q20908 PRELIMINARY; PRT; 770 AA.
AC Q20908;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE F56H9.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA BURTON J.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; Z74473; E249616; -.
SQ SEQUENCE 770 AA; 81625 MW; 771AB928 CRC32;

Query Match 10.4%; Score 448.5; DB 3; Length 770;
Best Local Similarity 23.7%; Pred. No. 3.4e-07;
Matches 183; Conservative 52; Mismatches 307; Indels 229; Gaps 22;

```
Qy     71 PGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTQAPT 130
Db     55 PSTTTSSATTTTVPSTTTSST-----TTTTVPPTTTSTTTTTTVPPTTTSTTTTTTVPPT 109

Qy    131 TTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTT 190
Db    110 TTSTTTTTTTVPPTTTSSTTTTTVPPTTTSSTTTTTVPPTTTTSTTTTTVPPTTTSTTTTT 169

Qy    191 QAPTTITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGC 250
Db    170 TVPATTTSTATTTTVPPTTSTTTTTTVPPTTTSTTTTTTVPPTTTSSTTTTTTTTLPTTT 229

Qy    251 PADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPA 310
Db    230 TT-----MMIP---ICSL-----DSLTF----GKGDNFQPQV------------------ 255

Qy    311 PPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRC--------------------D 349
Db    255 --------DIDVTYSNLVANIIPGTMDMQSLSTMQISCTAIAGYNAYMIFDVFNQPAENQ 306

Qy    350 KYYQCVHGNLVERR----------------------CGAGTHFSFEL------------ 375
Db    307 GFPQSVTVNAQCNRILINLFMFFKFLSYEKTSFYILCTSKTLSNFILLTVYFLYIGPTEL 366

Qy    375 --QQCDHIEL---------VGCTLPGGESEEVDVDEDACTGWYCPTE------------ 411
Db    367 LYTQIGSMNLNTYLNQTMGLGMYLQGPITQMIITINRFLVIWFTPTHVPQYSHRITLGAL 426

Qy    411 PIEWEP---------LPN--------------GCPADFSIDHLLPHESDCGQYLQCV--- 445
```

```
Db    427 SVSWITVTWLSTLIGLPGFDNNNNFDIIINLANCRVPIGFEHIGYYSTPCNNQITIIVVS 486

Qy    445 ----------------------------------HGQTIARPCPGNLHFSPATQSCESPV 470
                                             : : : :     ||    |
Db    487 GIFLIGFLTNFMNFMIGGKLIYTWKTLTIFRFLSNFKCVMKWISTLLHILLLT------- 540

Qy    471 TAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPT 530
              :: |  | | :| :||| ||   | | |       :  |   | | |    | |
Db    540 ------YKLDVINSCIAT-SPTADPTTTTTEATTTTEITTTTEEVTTTTEPVTTTTTPTT 592

Qy    531 TAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTA 590
          |:  :   |  ||   ||    :||| |   :   |  ||'|       | |:  || ||
Db    593 TSTTSTTTTTTPTTTTTT----TPTTTTSTTSTTTTTPTTTTSTTTTTTTTTTTATPTT 648

Qy    591 APTAAPTAAPTTAV--------PEIPTTVTSP---PTAAPTTAAPAPNTTVTVPPTAAPT 639
             |  |   |             |:|   ||      |  | |       |||
Db    649 TTTTMPPCNPNAVSLGMGDANNPQIFIDVTYNNYLSTTDPVTGIVTSTMTVTCSALNGYN 708

Qy    640 T--------AAPAPNTTVTVPPTAAPT--AAPPTVAHAPNTTAAPVTTTSA 680
          |          || |    :| | : |      |:      |   || | |
Db    709 TYMLFNGGQGGPADNQ--NLPQTISITLVCTSDTMVWNYEVTINGVTYTRA 757


RESULT   9
Q23635
ID   Q23635      PRELIMINARY;      PRT;   801 AA.
AC   Q23635;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO LONG TANDEM REPEAT REGION OF SIALIDASE.
GN   ZK84.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA   BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA   DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA   HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA   KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA   MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA   RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA   SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA   VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA   WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   KIRSTEN J.;
RL   SUBMITTED (APR-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U23181; G726436; -.
SQ   SEQUENCE   801 AA;  77123 MW;  7288A89D CRC32;


  Query Match             10.3%;  Score 447.5;  DB 3;  Length 801;
  Best Local Similarity   26.4%;  Pred. No. 3.8e-07;
  Matches  193;  Conservative   41;  Mismatches  288;  Indels  209;  Gaps   28;

Qy     10 ALGLVAARPE--VSDA-EKNPALHEPHPDCPPAEQHWLLPHEYDCTKFYYCEYGLKFIAP 66
          | | | :  |    || |  | : |     | ||:          |     |       |
Db    194 ASGYVTSEQEQGAGDAPEPAPVVEETPAPTPAAEE----------TPAPATEASNAVGTP 243

Qy     67 RDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTT-TTQAPTTT 125
             ||    :| : |          | | || | ||||:       |||   | || |
Db    244 EGYVDGTA---NAASAVAEPAVETPAPAPAAEETPAPATSSEEAPAPAPTAEETPAPETV 300

Qy    126 TQAP-----------TTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQ 173
          : ||            |  ||::  |     :   | ||     |::::    :|:
Db    301 SAAPEAANSYDSAGGDAATAPAPSSEAADAAAPTDSAASADTTAALVDTSSSEHAAESTE 360

Qy    174 APTTTQAPT-TTQAPTTT----------QAPTTITQAATTPAA-TTPAATTPAATTPAAT 221
          || |  | | || ||:             :|::| :    |||  |||  : ||    |
Db    361 APATDIAATETTPAPSVPAPVADAAAGYDSPSSIPEETPAPAAEDTPAPASAAAEETPAP 420

Qy    222 TPAATTPGVPAPTSAPVWPPICELLPNGCPADF----DIHLLIPHDKYCNLFYQCSNGYT 277
          |||     ||| :|   |   :      |||     |:    |                |
Db    421 APAAEE--TPAPETASAAP---DAAGGAAPADVAAPADVATTAPE-------------T 461

Qy    278 FEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPV---TEGNEDEDIDIGDLLDNGCPA 334
             :  | |  |       ||:     |:: || |   ||   |    ||       ||
Db    462 SSAQSAAGSYDVP-----SEPAS-----EVT-APIVESATEAPSDSAAPIGPAASEPAPA 510

Qy    335 NFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEE 394
                                                |    |   |     | |
Db    511 ----------------------------------PIEAPATDAATLETAPAPAAEPAP 534

Qy    395 VDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCP 454
             :|  |:  |:   |  | |
Db    535 A---AEAAAGYDAPSSVPEETPAP---------------------------------- 556

Qy    455 GNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPSTV 514
             :||     :|                    || |  | ||  |    |  ||:
Db    556 -----APAADETPAP-------------------APAAEETPAPAPAAEETPAPAPAAE 590

Qy    515 VPPATPPA---TAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTVTVPPTAAPTAAPTT 571
            ||| ||   | || |     | ||| |  | ||  |  |:|      |    | ||
Db    591 ETPATAPAAEETPAPAPAADETPAPAPAAEETPAPAPAVEETPAPA---PAVEETPAPAP 647

Qy    572 AVPEIPITVTSAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAPAPNTTVT 631
          |  | |     || ||  ||| |  | ||  |  | |    :|  ||  | ||||    |
Db    648 AAEETP---APAPAAAEYAAPVAEETPAPAPAAEETP----APAQAAEETPAPAPAAEET 700

Qy    632 VPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHAPNTTAAP----------------V 675
          | :|      |||       |  || ||| | | |    |||                 |
Db    701 -PASAPAAEETPAPAPAAEEAPAPAP-AAPETPAPAAEGAAAPAPIVSSGYDAAQSTDTV 758

Qy    676 TTTSAPATTPE 686
            :|||||: :
Db    759 VASSAPATSSD 769


RESULT  10
O76810
ID   O76810      PRELIMINARY;      PRT;   373 AA.
AC   O76810;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ICHIT PROTEIN.
GN   ICHIT.
OS   ANOPHELES GAMBIAE (AFRICAN MALARIA MOSQUITO).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; NEMATOCERA; CULICOIDEA; CULICIDAE; ANOPHELES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SUAKOKO; TISSUE=MAINLY MIDGUT;
RA   DIMOPOULOS G., SEELEY D., WOLF A., KAFATOS F.C.;
RT   "Malaria infection of the mosquito Anopheles gambiae activates immune
RT   responsive genes during critical transition stages of the parasite
RT   life cycle.";
RL   SUBMITTED (SEP-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AJ010903; E1320052; -.
SQ   SEQUENCE   373 AA;  40647 MW;  A7231CDB CRC32;


  Query Match             10.2%;  Score 442.5;  DB 3;  Length 373;
  Best Local Similarity   36.1%;  Pred. No. 3.1e-07;
  Matches  137;  Conservative   23;  Mismatches  124;  Indels   95;  Gaps   20;
```

```
Qy       1 MIKTLLFLTALGL----VAARPEVSDAEKNPALHEPHPDCPPAEQHWLLPHEY----DCT 52
           ||   :  ||||    |:|| |  :   |      ||:||  |    ||| :    :|:
Db       1 MIAAAMKYVALGLVLLAVSARAEPGEVIPN------HPNCP--EMQGPLPHYFIHPTNCS 52

Qy      53 KFYYC------EYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGC--TLPGPPAETTQAPA 104
           :|| |      ||        :|  |  |  :   |     | |    ||    ||  |
Db      53 RFYECHMRDAWEY--------ECPAGLHFNVAIDVCDFPVNAKCESQSPGDQTTTTLRPT 104

Qy     105 TTQ-APTTTQA----PTTTTQAPTTTTQAPTTTTQAPTTTQA---PTTTQA------PTT 150
           ||   ||||       ||||:| |||    |||| ||||       || |         |||
Db     105 TTTLRPTTTTTTDWITTTTTEATTTTKFPTTTTTSAPTTPSQWTDPTITTTTPVWTDPTT 164

Qy     151 TQAPTTTT------QAPTTTT------QAPTTTTQAPTTTQA--------PTTT------ 185
             ||||||        |||||         |||| |||||          ||||
Db     165 WSAPTTTTTWSDQPPPPTTTTTTVWTDPTATTTTHAPTTTTTWSDLPPPPPTTTTTVWID 224

Qy     185 -QAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATT--------PAATT--PGVPAP 233
            | ||| |||| |  :  |        ||   | |  ||         |  |:  |  | |
Db     225 PTATTTTHAPTTTTTWSDLPPPPPTTTTTTVWTDPTTTTTTDYTTAYPPTTSEPPSTPHP 284

Qy     234 TSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQ 293
           |  |  ||    |||              |  |: :| |  |   | :||:|||:|   :
Db     285 TD-PHCPPTGATLPN----------YWAHGTDCSRYYGCLEGCVKEFKCPDGLYWNDQQK 333

Qy     294 RCDSPANVECD-GEISPAP 311
           |||| :: :|   :| |||
Db     334 RCDSYSSSQCGCPDIPPAP 352
```

RESULT 11
P91365
ID P91365 PRELIMINARY; PRT; 2378 AA.
AC P91365;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CODED FOR BY C. ELEGANS CDNA YK50C7.5.
GN K06A9.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA GEISEL C., GATTUNG S.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; U80846; G1707247; -.
SQ SEQUENCE 2378 AA; 225242 MW; 765EDB23 CRC32;

Query Match 9.6%; Score 414.5; DB 3; Length 2378;
Best Local Similarity 26.3%; Pred. No. 7.1e-06;
Matches 158; Conservative 49; Mismatches 156; Indels 237; Gaps 27;

```
Qy     111 TTQAPTTTTQAPTTTTQAPTTTTQ-----APTTTQAPTTTQAPTTTQAPTTTTQAPTTTT 165
           ||  ||||  ||||    :|:||     :|||  || |  :|||  || |:    ||
Db     223 TTLGPTTTTITPTTTVVPASTSTQPAPASSPTTQPAPAT--SPTTQPAPATS----PTTQ 276

Qy     166 QAPTTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAATTPAATTPAATTPA- 225
           || |   :|||  || |  :|||  ||   |   | ||  |   | ||  |   | ||
Db     277 PAPAT---SPTTQPAPAT--SPTTQPAPA--TSPTTQPAPATSPTTQPAPATSPTTQPAP 329

Qy     225 ATTPGV-PAPTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCP 283
           ||:|   ||| ::|   |     |   |
Db     330 ATSPTTQPAPATSPTTQPTPATSPTTQP------------------------------- 358

Qy     284 EGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLP 343
                     :||        ||    |                 ||
Db     358 ------------TPAT-------SPTTQPT-----------------PA--------- 371

Qy     344 HGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPGGESEEVDVDEDACT 403

Db     371 ------------------------------------------------------------ 371

Qy     404 GWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLHFSPAT 463
                 | |   :| |   |                                 :| |     || |
Db     371 -----TSPTT-QPAPATSPT-------------------------TQPAPAT---SPTT 395

Qy     464 QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPATPPA- 523
           |   :| |:                |:  ||  :||  |  ||  :||  |:    | | ||
Db     396 Q--PAPATSP------------TTQPAPATSPTTQP--APATSPTTQPAPATSPTTQPAP 439

Qy     523 ----TAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTT-VTVPPTAAPTAAPTTAVPEIP 577
               | | |:    |||  :||  |: |   :|||  :|  :|:   || |:    |
Db     440 ATSPTTQPTPGTSPTTQPAPATSPTTQPSPAT--NPTTQPSVSSSASTQPAPATSGTTQP 497

Qy     578 ITVTSAPTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTAAPTTAAP--APNTTVTVPPT 635
             ||| |  |  | :|    || |:    |||  :| |:| | : |      ||:   ||
Db     498 APATSA-TTQPAPATSATTQPAPATS----PTTQPAPATSATTQSVPVVTTQTTMGTSPT 552

Qy     636 AAPTTAAPA---PNTTVTVPPTAAPTAAPPTVAHAPNTTAAPVTTTS-------APATTP 685
             |  :::||   |:| || | :   :::|  :  : :|:| |: ::|        | :|||
Db     553 TAGISSSPAATSPSTAVTSPSSLGTSSSP--LPSSISTSALPIASSSASSSPSAASSTTP 610
```

RESULT 12
Q17921
ID Q17921 PRELIMINARY; PRT; 825 AA.
AC Q17921;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE C12D12.1 PROTEIN.
GN C12D12.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,

```
RA   VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA   WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   NHAN M., HAWKINS J.;
RL   SUBMITTED (APR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U51998; G1255319; -.
SQ   SEQUENCE   825 AA;  86617 MW;  90FA8FCE CRC32;


  Query Match             9.4%;  Score 408;  DB 3;  Length 825;
  Best Local Similarity   24.0%;  Pred. No. 5.4e-06;
  Matches  148;  Conservative   26;  Mismatches  159;  Indels  284;  Gaps   17;

Qy    102 APATTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAP 161
          || ||  ||     | :| :| ||   ||| |  ||||  |    |  ||  |||   :
Db    460 APPTTPGPTFYPVTTVSTMSPPTTVTVPTTPTPVPTTTNTP---PANPTTATPTTVGTSK 516

Qy    162 TTTTQAP-----------TTTTQAPTTTQAPTTTQAPTTTQAPTTITQAATTPAATTPAA 210
           | | :|           :| | || |: :||||   ||   |||     | |  || :.
Db    517 QTNTISPHLSTITGSIVTSTPTMAPQTSASPTTTPTHTTASQPTT-----TKPVVTTNSV 571

Qy    211 TTPAATTPAATTPAATTPGVP-APTSAPVWPPICELLPNGCPADFDIHLLIPHDKYCNLF 269
          |    ||   | |  || | |   |:|||  |
Db    572 TPSTGTT---TVPVPTTTGSPTTQTTAPVTKP---------------------------- 601

Qy    270 YQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDEDIDIGDLLD 329

Db    601 ------------------------------------------------------------ 601

Qy    330 NGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVGCTLPG 389
                                                                    |:|
Db    601 ------------------------------------------------------------TVP- 604

Qy    390 GESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTI 449

Db    604 ------------------------------------------------------------ 604

Qy    450 ARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTA 509
                                                |:  || |   ||: |    |:
Db    604 -----------------------------------SSTTQTAPPVTTPTSQPPVTTTS 626

Qy    510 APSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTA-APESPTTVTVPPTAAPTAA 568
            :|:  | | | |   || :  :|| ||    || ||: ||     ||  || ||| :
Db    627 LLTTLTTP-TVPVTTTVVPSSATVPTTPPTTVTVAATTTSKAP----VVTTSPTLAPT-S 680

Qy    569 PTTAVPEIPITVTSAPTA-APTAAPTAAP------TAAPTTAVPEIPTTVTSPPTAAP-T 620
          ||       | || ::||| |    || ||      | |||   |  ||: |: ||  | |
Db    681 PTKLPTSPPSTVGTSPTAPANLTTPTTAPVNPTSSTTAPT--APVNPTSPTTAPTVPPVT 738

Qy    621 TAAPAPNTTVTVPPTAAPTTAAPAPNT-TVTVPPTAAPTAAPPT-VAHAPNTTAAPVTTT 678
          |  |   |: |   |   ||    | |  || | |  ||  | |  |  | | |  |  ||
Db    739 TTTPTTTTSTTTTTTTTTTTTQTTPTTPVVTTPSTITPTTRPVTQPASQPATQPATTPTT 798

Qy    679 SAPATTPEDDDIDPPLP 695
          : | ||     :   ||
Db    799 TIPTTTQFSSAVSVALP 815


RESULT  13
Q22579
ID   Q22579      PRELIMINARY;      PRT;  1777 AA.
AC   Q22579;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SIMILAR TO A. FAECALIS POLY(3-HYDROXYBUTYRATE) DEPOLYMERASE.
GN   T19D12.1.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA   BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA   DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA   HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA   KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA   MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA   RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA   SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA   VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA   WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   FAVELLO T.;
RL   SUBMITTED (DEC-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   WATERSTON R.;
RL   SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U41263; G1086828; -.
SQ   SEQUENCE   1777 AA;  179736 MW;  24064E32 CRC32;


  Query Match             9.1%;  Score 393;  DB 3;  Length 1777;
  Best Local Similarity   23.8%;  Pred. No. 2.5e-05;
  Matches  205;  Conservative   92;  Mismatches  283;  Indels  282;  Gaps   36;

Qy     62 KFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPP----------------AETTQAPAT 105
          | |   || | |    :  ||        |:| |                  |:|| ||
Db    187 KTIIQSDC-PST---VAPVTC-------STMPMPTTQPLVTSTAPEVTTTVAQTTTAPIV 235

Qy    106 TQAPTTTQAPTTT-------------TQAPTTTTQAPTTTTQAPTTTQAPTT-------- 145
          | | ||||  |||             | | |||  :  ||||  |||   ||
Db    236 TTANTTTQGVTTTAGVTTTVTRAQNSTLAATTTAPSTNTTTQGVTTTVGKTTTVTTAQNS 295

Qy    145 TQAPTTTQAPTTTTQAPTTTTQA---PTTTTQA-------PTTTQAPTTTQAPTTTQAPT 194
          | | ||| : |||    ||:|     ||| ||       |||||   ||| ||:|  |:
Db    296 TWAATTTASNTTTQPVVTTSTSTQGISTTTAQATPSSSVIPTTTQ---TTQRPTSTGIPS 352

Qy    195 TI---------------TQAATTPAAT-----TPAATTPAATTPAA-------------- 221
          |:                || :::  :|     ||: ||   |:  |
Db    353 TVSTSQGTSSTSPIPSTTQTSSSAPSTYTSNFTPSPTTTLLTSTIAPSTQGVPTSSKSSS 412

Qy    221 --TTPAAT-TPGVP-----------APTSAPVWPPICELLPNGCPADFDIHLLIPHDK- 265
            :||  | ||| |            : |  |  | :  |  :  |      |::
Db    413 PNSTPTTTITPGAPSSTLGSSSSTIVSTTITPSTPKVSTLTLSQSPTPTSTPLVVSSSSS 472

Qy    265 ------YCNLFYQCSNGY-TFEQRCPEGLYFNPYVQRCD---------------------- 297
                :    : |  |     |     ||  :
Db    473 GSSSTVVTSTITPSTQGVPTSTSNQPTPSTSNPTTPKSTVTASPSTTGATSTASPSTITS 532

Qy    297 ----------------------------------SPANVECDGEISPAPPVTEGNEDED 321
                                             :|    |   :: |  :| |  | |
Db    533 SAPTSQSHSPSSTMTSTVPVTSTFASTTTTVPITVAPGQCYCQSNVAVAFELTSGTSDLD 592

Qy    322 IDI-----------------GDLLDNGC-------PANFEIDWLLPHGNR--------- 348
          :||                 |||  ||       | :  :  |: :|
Db    593 LDIQNFIANYLFFYSGAPYILGDLTDNRTAISLVPFPNDDTLSDLMTYGAEQTPSGISAA 652

Qy    348 CDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIELVG----CTLPGGESEEVDVDEDACT 403
           | :    || |          |       :|  |      |     | |    |: |
```

```
Db    653 LDTFNILARGNAV---------ISDAFNIIPNITRKGYQGFVILVANSDESVQASVDSAT 703

Qy    404 -----GWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYLQCVHGQTIARPCPGNLH 458
Db    704 NLKAQGFNVITVAFKSSGKFDVLASQPSYNYTIVQDAD-KRYVATLIGNVLNTTYCNNLA 762

Qy    459 FSPATQSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPTAAPTAAPTAAPSTVVPPA 518
Db    763 ASSTT------MTSGVYSSTVTVTTQGSSSQAPSSTVT-VPTTGTTSG--AASTTGSITS 813

Qy    519 TPPATAAPVPPTTAIPTPAPTAAP---TAAPTTAAPESPTTVTVPPTAAPTAAPTTAVPE 575
Db    814 TQQATS-----TSSVITTGSTSAPQSSTAVSSTTTSPSTTTGSTPAPQSSTVASTTTVS- 868

Qy    576 IPITVTS-----------------APTAAPTAAPTAAPTAAPTTAVPEIPTTVTSPPTA 617
Db    868 -PYTTTECICTTVSNTFGTSTTQGLTSSSAQSTASTGVSTVASSTTIPQ-GSSSSSP--Q 923

Qy    618 APTTAAPAPNT-----TVTVPPTAAPT------TAAPAPNTTVTVPPTAAPTAAPPTVAH 666
Db    924 SPTSAQPAQSSSTSAATVTVSSSQSPTSSPAQQSSTPAGSSTVTVQQSSSFQSPQSTTQI 983

Qy    667 APNTTA----APVTTTSAPATT 684
Db    984 GSSTTVPSTQAPSSTSGGPTTT 1005


RESULT  14
Q41805
ID   Q41805      PRELIMINARY;      PRT;  1188 AA.
AC   Q41805;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   EXTENSIN-LIKE PROTEIN PRECURSOR.
OS   ZEA MAYS (MAIZE).
OC   EUKARYOTA; VIRIDIPLANTAE; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;
OC   TRACHEOPHYTA; EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA;
OC   POALES; POACEAE; ZEA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B73; TISSUE=POLLEN;
RA   RUBINSTEIN A.L., BROADWATER A.H., LOWREY K., BEDINGER P.A.;
RL   SUBMITTED (JUN-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; Z34465; G600118; -.
DR   PFAM; PF00560; LRR; 3.
KW   SIGNAL.
FT   SIGNAL        1     27       POTENTIAL.
SQ   SEQUENCE   1188 AA;  120980 MW;  47E3D9BA CRC32;


  Query Match             9.1%;  Score 392;  DB 8;  Length 1188;
  Best Local Similarity   22.5%;  Pred. No. 2e-05;
  Matches  170;  Conservative   87;  Mismatches  324;  Indels  176;  Gaps   28;

Qy      2 IKTLLFLTALGLVAARPEVSDAEKNPALHEPHPDC-----PPAEQHWLLPHEYDCTKFYY 56
Db    520 VKTTSPPAPIGSPSPPPPVSVVSPPPPVKSPPPPAPVGSPPPPEKSPPPP---------- 570

Qy     57 CEYGLKFIAPRDCAPGTEFKFSAQTCVHAALAGCTLPGPPAETTQAPATTQAP------- 110
Db    570 --------APVASPPPPVKSPPPPTLVASPPPPVKSPPPPAPVASPPPPVKSPPPPTPVA 621

Qy    110 -TTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAP 168
Db    622 SPPPPAPVASSPPPMKSPPPPTPVSSPPPPEKSPPPPPPAKSTPPPEEYPTPPTSVKSSP 681

Qy    169 TTTTQAPTTTQAP------------TTTQAPTTTQAPTTITQAATTPAATTPAATTPA-- 215
Db    682 PPEKSLPPPTLIPSPPPQEKPTPPSTPSKPPSSPEKPSPPKEPVSSPPQTPKSSPPPAPV 741

Qy    215 -------ATTPAATTPAATTPGV-----PAPTSAPVWPPICELLPNGCPADFDIHLLIPH 262
Db    742 SSPPPTPVSSPPALAPVSSPPSVKSSPPPAPLSSP--PPAPQV--KSSPPPVQVSSPPPA 797

Qy    263 DKYCNLFYQCSNGYTFEQRC-PEGLYFNPYVQRCDSPANVECDGEISPAPPVTEGNEDED 321
Db    798 PKSSPPLAPVSSPPQVEKTSPPPAPLSSPPLAPKSSPPHV----VVSSPPPVVKSSPP-- 852

Qy    322 IDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHFSFELQQCDHIE 381
Db    852 -----------PAPVSSPPLTP-----------------KPASPPAHVS------SPPE 876

Qy    382 LVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHLLPHESDCGQYL 441
Db    877 VVKPSTPPAPTTVISP----------PSEPKSSPP-----PTPVSLPPPIVKSS------ 916

Qy    442 QCVHGQTIARPCPGNLHFSPAT-QSCESPVTAGCQVFECDSDNQCTSTAAPTAAPTAAPT 500
Db    916 ----------PPPAMVSSPPMTPKSSPPPVVV--------SSPPPTVKSSPPPAPVSSPP 957

Qy    501 AAPTAAPTAAPSTVVPP---ATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAAPESPTTV 557
Db    958 ATPKSSPPPAPVNLPPPEVKSSPPPTPVSSPP------PAPKSSPPPAPMSSPP--PPEV 1009

Qy    558 TVPPTAAPTAAPTTAV--PEIPITVTSAP---TAAPTAAPTAAP-----TAAPTTAVPEI 607
Db   1010 KSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPISSP 1069

Qy    608 PTTVTSPPTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTAAPTAAPPTVAHA 667
Db   1070 PPPVKSPPPPAPVSSPPPP---VKSPPPPAPVSSPPPP---IKSPPPPAPVSSPPPAPVK 1123

Qy    668 PNTTAAPVTTTS-APATTP-----EDDDIDPPLPNDP 698
Db   1124 PPSLPPPAPVSSPPPVVTPAPPKKEEQSLPPPAESQP 1160


RESULT  15
O39307
ID   O39307      PRELIMINARY;      PRT;   750 AA.
AC   O39307;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   POSITIONAL COUNTERPART OF HSV-1 GENE US5.
GN   71.
OS   EQUINE HERPESVIRUS 4.
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC   ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN   [1]
RP   SEQUENCE OF 685-750 FROM N.A.
RC   STRAIN=NS80567;
RX   MEDLINE; 93389454.
RA   CULLINANE A.A., NEILAN J., WILSON L., DAVISON A.J., ALLEN G.;
RT   "The DNA sequence of the equine herpesvirus 4 gene encoding
RT   glycoprotein gp17/18, the homologue of herpes simplex virus
RT   glycoprotein gD.";
RL   J. GEN. VIROL. 74:0-0(0).
RN   [2]
RP   SEQUENCE OF 1-111 FROM N.A.
RC   STRAIN=NS80567;
RX   MEDLINE; 93119267.
RA   NAGESHA H.S., CRABB B.S., STUDDERT M.J.;
RT   "Analysis of the nucleotide sequence of five genes at the left end of
RT   the unique short region of the equine herpesvirus 4 genome.";
RL   ARCH. VIROL. 128:143-154(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NS80567;
RX   MEDLINE; 98264497.
RA   TELFORD E.A.R., WATSON M.S., PERRY J., CULLINANE A.A., DAVISON A.J.;
RT   "The DNA sequence of equine herpesvirus-4.";
RL   J. GEN. VIROL. 79:1197-1203(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=NS80567;
RA   TELFORD E.A., WATSON M.S., PERRY J., CULLINANE A.A., DAVISON A.J.;
RL   SUBMITTED (OCT-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF030027; G2606019; -.
SQ   SEQUENCE   750 AA;  77206 MW;  7407BDC3 CRC32;


  Query Match             9.0%;  Score 388;  DB 11;  Length 750;
  Best Local Similarity   24.0%;  Pred. No. 1.9e-05;
  Matches  153;  Conservative   45;  Mismatches  168;  Indels  272;  Gaps   13;

Qy      78 SAQTCVHAALAGCTLPGPPAETTQAPATTQAPTTTQAPTTTTQAPTTTTQAPTT---TTQ 134
           || :  ::: :  | |   :  | :  :|| |:|:| |||::   :||| : :|   :|
Db      40 SASSETNSSSSPTTGPTTTSSQTSSSNSTQTPSTSQTPTTSSSTVSTTTTSNSTNESSTA 99

Qy     135 APTTTQAPTTTQAPTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTQAPTT--TQAPTTTQA 192
             |:|  ||:|:| |:|   |: :::||:|    | || | |||:: ||  | | ||| |
Db     100 TATSTATPTSTEASTSTTTSTSVSESPTST----TATTAATTTTESTTTESTTAATTTAA 155

Qy     193 PTT--ITQAATTPAATTPAATTPAATTPAATTPAATTPGVPAPTSAPVWPPICELLPNGC 250
            ||   | |||| |||| |||| |||| |||| ||||    | |:|
Db     156 TTTAATTTAATTTAATTTAATTTAATTTAATTTAATT--TAATTTA-------------- 200

Qy     251 PADFDIHLLIPHDKYCNLFYQCSNGYTFEQRCPEGLYFNPYVQRCDSPANVECDGEISPA 310

Db     200 ------------------------------------------------------------ 200

Qy     311 PPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERRCGAGTHF 370

Db     200 ------------------------------------------------------------ 200

Qy     371 SFELQQCDHIELVGCTLPGGESEEVDVDEDACTGWYCPTEPIEWEPLPNGCPADFSIDHL 430

Db     200 ------------------------------------------------------------ 200

Qy     431 LPHESDCGQYLQCVHGQTIARPCPGNLHFSPATQSCESPVTAGCQVFECDSDNQCTSTAA 490
                                          || :  :   |             |:|||
Db     200 -------------------------------ATTTAATTTAA------------TTTAA 215

Qy     491 PTAAPTAAPTAAPTAAPTAAPSTVVPPATPPATAAPVPPTTAIPTPAPTAAPTAAPTTAA 550
            | | |   ||| | | |   :|     |  ||      | | || :  |: | | |||
Db     216 TTTAATT--TAATTTAATTTAATTTAATTTAATTTAATTTAATPTESSEAASSTLAATTAD 273

Qy     551 PESPTTV-TVPPTAAPTAAPTTAVPEIPITVTSAPTAAPTAAPTAAP-TAAPTTAVPEIP 608
             : ||  |   | | | | |||      | ||  ||| | : |:|  | ||||   :
Db     274 TTADTTADTTADTTADTTADTTADTTTTSGSTAANTTSTTSATVTIAPTTFTTKYT 333

Qy     609 TTVTSP---PTAAPTTAAPAPNTTVTVPPTAAPTTAAPAPNTTVTVPPTA-APTAAPPT- 664
           |  :|     |:  |   |   |  | || | :  |       | |||   ||   ||
Db     334 TNSSSTGKINTSKNTPKPPQYTTASTEKPTKANSLTAANATGLSTKPPTLFTPTQTSPTP 393

Qy     664 ------------VAHAPNTTAAPVTTTSAPATTPEDDD 689
                       : :   |   |    |:    |  |:
Db     394 SETSVGTREYLAITYGKTTYTTPTNALSSTNVWPARDN 431


Search completed: July 18, 1999, 02:42:01
Job time: 3624 sec
```